AD-A254 615

ION PAGE

Form Approved
OMB No. 0704-0188

1. A

3. REPORT TYPE AND DATES COVERED
ANNUAL 01 Feb 91 TO 31 Jan 92

4. TITLE AND SUBTITLE
VISUAL MOTION PERCEPTION AND VISUAL INFORMATION PROCESSING

5. FUNDING NUMBERS
G AFOSR-91-0178
PE 61102F
PR 2313
TA A5

(2)

6. AUTHOR(S)
Dr George Sperling

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Dept of Psychology
New York University
6 Washington Place, RM 980
New York, NY 10003

8. PERFORMING ORGANIZATION REPORT NUMBER
AFOSR-TR- 92 0760

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Dr John F. Tangney
AFOSR/NL
Building 410
Bolling AFB DC 20332-6448

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

DTIC
ELECTE
AUG 18 1992
S A D

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release;
distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The publications describe progress in two related areas of visual information processing: motion processing and visual attention. The full equivalence between Reichart motion detection and Fourier motion analysis (first-order motion processing) was proved formally. A new experimental paradigm was developed to test the model of nonFourier (2nd-order) motion processing. This model, which accounts for the perception of motion-from-texture, consists of a stage of linear spatio-temporal filtering followed by fullwave rectification and then by standard (Reichart) motion analysis. It was demonstrated that human 2nd-order motion is, for practical purposes, one-dimensional (i.e., a single channel system). The spatial filter that this channel utilizes was measured and found to be lowpass. Work on attentional processes in visual task using rapid sequences of superimposed patterns showed that highly trained subjects were unable to use gross physical differences to filter out unattended items at an early stage of perceptual processing. On the contrary, the results are explained by postulating that attended and unattended elements of the input are tagged as such at an early stage, and are then discriminated later on the basis of this tag.

14. SUBJECT TERMS

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
(U)	(U)	(U)	(U)

USAF Office of Scientific Research, Life Sciences Directorate, Visual Information Processing Program

Interim Progress Report, 01 Feb 1991 - 31 Jan 1992

Grant AFOSR 91-0178

Visual Motion Perception

George Sperling, New York University

ABSTRACT

The reports enclosed with this report describe experiments related to four aspects of visual information processing: The main thrust is continuing studies of two separate motion-computation systems and the derivation of the function properties of each. The pronoun we is used in this report to refer to the PI in conjunction with one or more of the other investigators, students, and staff.

(1) The most significant new work is described in a published abstract and a preprint by the PI with Peter Werkhoven and Charles Chubb. Using a new paradigm (experimental display plus analysis), it was found that second-order motion perception for locally parallel textures is quite well approximated by a single-channel system. Previous studies (by other authors) that asserted otherwise were shown to have contained incorrect analyses. Elaborations of this paradigm (now in progress) will enable us to establish the full dimensionality of motion and of texture processing (analogously to the dimensionality of color vision). The manuscript has been accepted for publication in Vision Research, pending optional revisions, which are in progress.

(2) A paper describing the formal proof of the equivalence of Reichart detectors and Fourier analysis (of motion and texture) stimuli was published by the Journal of Mathematical Psychology. The paper also contains three illustrative experiments on texture-from-motion, the last of which demonstrates that the rectifying nonlinearity cannot be a pure square function. that

(3) A paper (Sutter, Sperling, & Chubb) describing research that enabled the determination of the partial selectivity of second-order pattern perceivers was completed and submitted to Vision Research for publication.

(4) Studies of the detection and discrimination of visual acceleration. These two papers represent work that Werkhoven continued with Dutch collaborators during his period at NYU. Just as motion-from-texture involves the analysis of spatio-temporal modulation in texture, the detection of acceleration involves spatio-temporal modulations in velocity. Werkhoven, Snippe and Toet ingeniously extend the principles that have been used in other studies of second-order perception to derive a model of acceleration detection based on a linear systems analysis of velocity variation. Snippe and Werkhoven apply a similar model to account for the detection of pulse modulations of velocity.

(5) The mechanism of nonspatial attentional selection. A manuscript describing a repetition detection paradigm developed by the PI in collaboration with Steve Wurst was completed and accepted for publication. A rapid sequence of 30 stimuli occurs at a single location. The subject must detect an embedded repetition. Successive items alternate in a particular feature value (e.g., black items versus white items on gray), and the subject is instructed to attend only to one value of the feature (e.g., white). The main result is

that even unattended items enter memory. A theory account for many complex and paradoxical results is that attention acts as a feature, e.g. A+ for an attended item, A- for an unattended item. Subsequent processes treat this top-down "attention feature" just as if were another stimulus feature.

(6) Work in progress is sketched briefly under "Students."

The main activities throughout this grant have been carrying out the experimental research set forth in the proposal (1990), following up promising leads that developed in the course of this work, and preparing manuscripts for publication. The work is best described by the publications and technical reports; these are appended. An overview, including facilities and personnel, is provided below.

FACILITIES

The Human Information Processing Laboratory (HIPL) is highly versatile laboratory for conducting research in almost any area of vision or cognition as described in previous progress reports and the current proposal.

PERSONNEL

Principle investigator. George Sperling, Professor of Psychology and Director of the Human Information Processing Laboratory. As projected in the original proposal, the PI devoted 10% time during 9 month academic year plus 50% time during 3 summer months totaling 26.67% of full time averaged over the full year)

Full-time

Research Associate, Dr. Peter Wenhoven worked primarily on visual motion and on related mathematical issues.

Systems Programmer. David Tanzer, a PhD student in Computer Science at NYU's Courant Institute is being employed full-time as a systems programmer. He is experienced, highly skilled, and effective. Beginning in September, 1991, NYU contributed 1/2 of Tanzer's salary.

Part-time

Consultant. Dr. Barbara Dosher. During this period, Dr. Dosher collaborate in preparing previously executed projects for publication (6 days).

Administrative assistant. Ms. Pamela Stark, a graduate student in the Department of Applied Science, Ms. Stark took an indefinite pregnancy leave just prior to the end of the current period. After a period of search, she was replaced by Paula Azevedo.

Graduate students

Joshua Solomon. Beginning his final year at NYU, Solomon has been and continues to work on three projects in visual psychophysics: (1) the lateral inhibition of apparent contrast by adjacent fields of high contrast; (2) discriminating half-wave and full-wave mechanisms of second-order motion and texture detection; and (3) the peripheral visibility of second-order motion and texture displays.

Shui-I Shi. Ms. Shi has been working on information processing studies to test attentional theories. The main project involves a unified attention theory to account for attention gating experiments and iconic memory--the link between attentional gating and reaction time studies of attention having been previously established by Erich

Weichselgartner in the HIPL. These are empirical studies of attention plus extensive Monte Carlo simulations of a comprehensive model. Additionally, Ms. Shi is extending the methods to a study of attentional control of visual search.

HIP Lab Publications, 1991

1991 Landy, Michael S., Barbara A. Dosher, George Sperling, and Mark E. Perkins. Kinetic depth effect and optic flow: 2. Fourier and non-Fourier motion. *Vision Research*, 1991, *31*, 859-876.

1991 Parish, David H. and George Sperling, Object spatial frequencies, retinal spatial frequencies, noise, and the efficiency of letter discrimination. Vision Research, 1991, *31*, 1399-1415.

1991 Solomon, Joshua A, and George Sperling. Can we see 2nd-order motion and texture in the periphery? *Investigative Ophthalmology and Visual Science*, ARVO Supplement, 1991, *32, No. 4*, 714. (Abstract)

1991 Werkhoven, Peter, Charles Chubb, and George Sperling (1992). Testure-defined motion is ruled by an activity metric--not by similarity. *Investigative Ophthalmology and Visual Science*, ARVO Supplement, 1991, *32, No. 4*, 829. (Abstract)

1991 Sutter, Anne, George Sperling and Charles Chubb, Further measurements of the spatial frequency selectivity of second-order texture meachanisms. *Investigative Ophthalmology and Visual Science*, ARVO Supplement, 1991, *32, No. 4*, 1039. (Abstract)

1991 Chubb, Charles, and George Sperling. Texture quilts: Basic tools for studying motion-from-texture. *Journal of Mathematical Psychology*, 1991, *35*, 411-442.

1991 Chubb, Charles, Joshua A. Solomon, and George Sperling. Contrast contrast determines perceived contrast. *Optical Society of America Annual Meeting Technical Digest, 1991*, Vol. 17. Washington D.C.: Optical Society of America, 1991. P. XX. (Abstract)

1991 Sperling, G. and Wurst, S. A. (1991). Selective attention to an item is stored as a feature of the item. *Bulletin of the Psychonomic Society*, 1991, *29*, XX. (Abstract)

Papers Under Submission for Publication, Technical Reports

1991 Sperling, G. and Wurst, S. A. (1992). Using repetition detection to define and localize the processes of selective attention. In D. E. Meyer and S. Kornblum (Eds.), *Attention and Performance XIV: Attention and Performance XIV: Synergies in Experimental Psychology, Artificial Intelligence, and Cognitive Neuroscience - A Silver Jubilee* Cambridge, MA: MIT Press (In press.)

1991 Werkhoven, Peter, George Sperling, and Charles Chubb (1992). Motion perception between dissimilar gratings: A single channel theory. Vision Research, 1992, *32*. (In press.)

1991 Werkhoven, P., Snippe, H. P., and Toet, A. (1991). Visual processing of optic acceleration. Submitted to Vision Research.

1991 Snippe, H. P., and Werkhoven, P. (1991). Pulse modulation detection in human motion vision. Submitted to Vision Research.

Invited Lectures at Universities and Institutes

1991 Department of Psychology Colloquium, University of California, Irvine, Irvine, CA, January 10, 1991. *Visual Preprocessing.*

1991 Department of Psychology University of California at San Diego, La Jolla, CA, February 28, 1991. *Mechanisms of Attention.*

1991 University of California, Berkeley Berkeley, California, Joint Cognitive Science Colloquium and Oxyopia Colloquium (Optometry School), March 22, 1991. *Visual Preprocessing.*

1991 University of California, Berkeley Berkeley, California, Department of Psychology/Cognitive Science Colloquium, March 22, 1991. *The Spatial, Temporal, and Featural Mechanisms of Visual Attention.*

1991 Bonny Center for the Neurobiology of Learning and Memory, University of California, Irvine, Irvine, CA, April 8, 1991. *Mechanisms of Visual Attention.*

1991 Salk Institute, University of California at San Diego, La Jolla, CA, April 10, 1991. *Visual Preprocessing.*

1991 Department of Psychology, University of Florida at Gainsville, April 26, 1991. *Systems and Stages of Visual Processing.*

1991 Shanghai Institute of Technical Physics, Shangahi, China, June 17, 1991. *How the Human Visual System Computes Visual Motion* [Host: Prof. Kuang, Ding Bo (Director, SITP); Translators: Dr. Zhang, Ming and Chen, Lulin.]

1991 Department of Computer Science, Shanghai Information-Technology Engineers Examination Center, Fudan University, Shangahi, China, June 18, 1991. *Neural Principles of Preprocessing for Human Pattern Recognition.* [Host: Prof. Wu, Lide (Director, SITEEC).]

1991 Department of Electronic Science and Technology, Institute of Applied Electronics, East China Normal University, Shangahi, China, June 20, 1991. *Measuring Attention* and *How the Human Visual System Computes Visual Motion* [Host: Prof. Weng, Moying (Chairman and Director); Translator: Dr. Zhang, Ming.]

1991 Department of Psychology, Beijing University, and Institute of Psychology, Chinese Academy of Sciences, Beijing, China, June 25, 1991. [Host: Prof. Jing, Qicheng (Director, Institute of Psychology)]

Morning: *The Efficiency of Pereception* [Translators: Dr. Zhang, Ken and Prof. Jing, Qicheng.]

Afternoon: *Measuring Attention.* [Translator: Luo, Chun-Rong.]

1991 Computational Vision Laboratory, Institute of Biophysics, Chinese Academy of Sciences, Beijing, China, June 28, 1991. *First- and Second-Order Motion Perception.* [Host: Prof. Wang Shuo-Rong (Director, Institute of Biophysics); Translator: Prof. Wang, Yun-Jiu (Laboratory Director.]

1991 New York University, Cognitive Sciences Colloquium, September 12, 1991. *Is There Attentional Filtering of Items by Feature as Well as by Location?*

Joshua A. Solomon and George Sperling. Can We See 2nd-Order Motion and Texture in the Periphery? Investigative Opthalmology and Visual Science, 1991, **32, No. 4**, *ARVO Supplement,* 714

CAN WE SEE 2nd-ORDER MOTION AND TEXTURE IN THE PERIPHERY?

Joshua A. Solomon and *George Sperling*.

Human Information Processing Laboratory, New York University

Stimuli. Our 1st-order stimuli are moving sine gratings. Our 2nd-order stimuli are patches of static visual noise, whose contrasts are modulated by moving sine gratings. Neither the spatial orientation nor the direction of motion of these 2nd-order (drift-balanced) stimuli can be detected by analysis of their Fourier domain power spectra. They are invisible to Reichardt and motion-energy detectors.

Method. For these dynamic stimuli, in the fovea, and at 12 deg eccentricity, we measured contrast modulation thresholds as a function of spatial frequency for discrimination of ± 45 deg texture slant and for discrimination of direction of motion. Spatial frequency was varied by changing viewing distance.

Results. For sufficiently low spatial frequencies and sufficiently large contrast modulations, all stimuli are visible both foveally and peripherally. For peripherally viewed 1st-order gratings, the highest spatial frequency at which motion or texture discrimination is possible is about 1/4 that at which the corresponding discrimination is possible for foveally viewed gratings. For peripherally viewed 2nd-order gratings, the highest spatial frequencies at which motion or texture discrimination are possible are somewhat less than 1/4 the frequencies of the corresponding foveal discriminations. Thus, as the stimulus moves peripherally, the visual mechanisms that detect 2nd-order motion and texture lose sensitivity somewhat faster than the 1st-order mechanisms.

Conclusions. Under certain specific assumptions, our results suggest the following about the neural detectors involved in these discriminations: (1) For both motion and texture, there are more foveal than peripheral detectors at all spatial frequencies. (2) There are more 1st-order than 2nd-order detectors. (3) On the average, foveal detectors respond to higher spatial frequencies than peripheral detectors. (4) The 2nd-order foveal–peripheral spatial frequency difference is somewhat larger than the 1st-order difference.

Supported by AFOSR Life Sciences, Visual Information Processing Program, Grant 88-0140.

Peter Werkhoven, Charles Chubb and George Sperling. Texture-Defined Motion is Ruled by an Activity Metric--Not by Similarity. Investigative Opthalmology and Visual Science, 1991, **32, No. 4**, *ARVO Supplement,* 829

TEXTURE-DEFINED MOTION IS RULED BY AN ACTIVITY METRIC – NOT BY SIMILARITY

Peter Werkhoven, Charles Chubb and *George Sperling*.

Human Information Processing Laboratory, New York University

We examined motion carried by textural properties. The stimuli we used consisted of patches of sinusoidal grating of various spatial frequencies and contrasts. Phases were randomized to insure that motion mechanisms sensitive to correspondences in stimulus luminance were not systematically engaged.

We used an ambiguous apparent motion paradigm in which a "heterogeneous" motion path (defined by alternating patches of a type A and a type B texture) competes with a "homogeneous" motion path defined by patches of type A. We found that the strength of these (2nd order) motion stimuli is determined by the covariance of the *activity* of the textures that define the motion paths. The activity of a texture is an hypothesized property that is proportional to the texture's contrast and is found to be inversely proportional to its spatial frequency (within the range of spatial frequencies examined). Indeed, heterogeneous motion between equal contrast patches of a high spatial-frequency texture A and a low-spatial frequency texture B can easily dominate homogeneous motion between two patches of A because the activity of texture B is higher than that of texture A.

At temporal frequencies higher than 4 Hz, we find that activity covariance almost exclusively determines motion strength. At lower temporal frequencies, similarity between textures becomes a significant factor as well.

Supported by AFOSR Life Sciences, Visual Information Processing Program, Gran: 88-0140

Anne Sutter, George Sperling and Charles Chubb. Furthur Measurements of the Spatial Frequency Selectivity of Second-Order Texture Mechanisms. Investigative Opthalmology and Visual Science, 1991, **32, No. 4**, *ARVO Supplement*, 1039

FURTHER MEASUREMENTS OF THE SPATIAL FREQUENCY SELECTIVITY OF SECOND-ORDER TEXTURE MECHANISMS

Anne Sutter, George Sperling, & Charles Chubb

Human Information Processing Laboratory, New York University, NY, NY 10003

A number of investigations of texture and motion perception suggest a two-stage processing system consisting of an initial stage of selective linear filtering, followed by a rectification and a second stage of selective linear filtering. Here we present new data measuring two properties of the second-stage filters: their contrast modulation sensitivity as a function of spatial frequency (MTF), and the relation of initial spatial filtering to second-stage selectivity. To determine the MTF, we used a staircase procedure to obtain amplitude modulation thresholds for the detection of the orientation of Gabor modulations of a bandlimited noise carrier. We used improved noise carriers with a narrower bandwidth than the stimuli reported last year. Four carrier bands were created with center frequencies of 2, 4, 8, and 16 c/deg. The spatial frequency of the test signals (Gabor amplitude modulations) ranged from 0.5 to 8 c/deg.

The improvements in our stimuli produced a different pattern of results: (1) The threshold amplitude of signal modulation was lowest for 0.5 and 1.0 c/deg. Above 1.0 c/deg, threshold increased with frequency[1]. (2) There was a significant interaction of carrier frequency band with the modulating frequency, with the lowest thresholds occuring for carrier frequency/modulation frequency ratios of about three to four octaves. These results indicate that the second-stage selective filters and detectors are most sensitive to frequencies lower than or equal to 1 c/deg, and that they are selective with regard to the spatial frequency content of the carrier noise on which the signals are impressed.

[1]Jamar, J.H.T. & Koenderink, J.J., (1985). *Vis. Res.* 25 (4) pp. 511-521.

Supported by AFOSR Life Sciences Directorate Grant 88-0140 and NIMH Grant 5T32MH14267.

Reprinted from JOURNAL OF MATHEMATICAL PSYCHOLOGY
Vol. 35, No. 4, December 1991
Printed in Belgium

Texture Quilts: Basic Tools for Studying Motion-from-Texture

CHARLES CHUBB

Department of Psychology, Rutgers University

AND

GEORGE SPERLING

Psychology Department and Center for Neural Sciences, New York University

A theoretical foundation and concrete stimulus-construction methods are provided for studying motion-from-spatial-texture without contamination by motion mechanisms sensitive to other aspects of the signal. Specifically, examples are constructed of a special class of random stimuli called *texture quilts*. Although, as we demonstrate experimentally, certain texture quilts display consistent apparent motion, it is proven that their motion content (a) is unavailable to standard motion analysis (such as might be accomplished by an Adelson Bergen motion-energy analyzer, a Watson Ahumada motion sensor, or by any elaborated Reichardt detector), and (b) cannot be exposed to standard motion analysis by any purely temporal signal transformation no matter how nonlinear (e.g., temporal differentiation followed by rectification). Applying such a purely temporal transformation to any texture quilt produces a spatiotemporal function P whose motion is unavailable to standard motion analysis: The expected response of every Reichardt detector to P is 0 at every instant in time. The simplest mechanism sufficient to sense the motion exhibited by texture quilts consists of three successive stages: (i) a purely spatial linear filter, (ii) a rectifier (but not a perfect square law) to transform regions of large negative or positive responses into regions of high positive values, and (iii) standard motion analysis.

1. INTRODUCTION

Standard Motion Analysis. The extensive literature on the motion of random-dot cinematograms (Anstis, 1970; Baker & Braddick, 1982a, 1982b; Bell & Lappin, 1979; Braddick, 1973, 1974; Chang & Julesz, 1983a, 1983b, 1985; van Doorn & Koenderink, 1984; Julesz, 1971; Lappin & Bell, 1972; Nakayama & Silverman, 1984; Ramachandran & Anstis, 1983) points toward the view that a "short-range" system (Braddick, 1973, 1974) submits the raw spatiotemporal luminance function directly to *standard motion analysis* (such as might be accomplished by an Adelson Bergen motion-energy detector (Adelson & Bergen, 1985), a Watson Ahumada

Reprint requests should be sent to Charles Chubb, Department of Psychology, Rutgers University, New Brunswick, NJ 08903 or George Sperling, HIP Lab, NYU, 6 Washington Place, New York, NY 10003.

0022-2496/91 $3.00

motion sensor (Watson & Ahumada, 1983a, 1983b, 1985), an elaborated Reichardt detector (van Santen & Sperling, 1984, 1985), or some variants of a gradient detector (Marr & Ullman, 1981; Adelson & Bergen, 1986).

Fourier and Non-Fourier Mechanisms. An impressive number of observations suggest that standard motion analysis is not the whole story (Bowne, McKee, & Glaser, 1989; Cavanagh, Arguin, & von Grunau, 1989; Derrington & Badcock, 1985; Derrington & Henning, 1987; Green, 1986; Lelkins & Koenderink, 1984; Pantle & Turano, 1986; Petersik, Hicks, & Pantle, 1978; Ramachandran, Ginsburg, & Anstis, 1983; Ramachandran, Rao, & Vidyasagar, 1973; Sperling, 1976; Turano & Pantle, 1989). In particular, Chubb and Sperling (1987, 1988) have demonstrated a variety of stimuli that display consistent, unambiguous apparent motion, yet that do not systematically stimulate mechanisms that apply standard motion analysis directly to luminance. For reasons that become clear in Section 2, we call any motion system that applies standard analysis to the raw signal as a *Fourier* mechanism, and we refer to any system that applies standard analysis to a nonlinear transformation of the signal as a *non-Fourier* mechanism.

Microbalanced Stimuli. The methods used by Chubb & Sperling to construct stimuli whose obvious and consistent motion content cannot be revealed by applying standard motion analysis directly to luminance are founded on the notion of a *microbalanced* random stimulus. In Section 2.3.5, we show that the expected response of any standard motion analyzer applied directly to any microbalanced random stimulus is equal to the expected response of the corresponding analyzer tuned to motion of the same type, but in the opposite direction.

Microbalanced random stimuli allow us to differentially stimulate non-Fourier motion mechanisms without systematically engaging Fourier mechanisms. This is the source of their importance in the study of motion perception.

There are probably several types of non-Fourier motion mechanisms, distinguished by the different transformations they apply to the signal prior to standard motion analysis. In this paper, we extend the theory of microbalanced random stimuli in order to develop methods for constructing stimuli that selectively engage specific classes of non-Fourier mechanisms without stimulating either Fourier mechanisms or other classes of non-Fourier mechanisms.

Pointwise Transformations; Static Nonlinearities. A transformation T is called *pointwise* if the output of T at any point (x, y, t) in space-time depends only on the (stimulus) input value at that point. A *nonlinear* pointwise transformation sometimes is called a *static nonlinearity*. For instance, simple rectifiers and thresholders are pointwise transformations. In Section 3, we address the problem of isolating the class of non-Fourier mechanisms that apply a simple pointwise transformation prior to standard motion analysis from the class of all those mechanisms that apply more complicated transformations. The central result in this section is proposition 3.2 which provides necessary and sufficient conditions for a random stimulus I to be such that any pointwise transformation of I is microbalanced.

Purely Temporal Transformations and Texture Quilts. The results with pointwise transformations are extended in Section 4 to purely temporal transformations (defined in Section 2.2). Whereas, for a pointwise transformation, the transformed value at the point (x, y, t) depends only on the stimulus value at (x, y, t), in a purely temporal transformation the transformed value at (x, y, t) may depend in any way whatsoever on the entire history of stimulus values at (x, y). We define the class of stimuli called *texture quilts* (Definition 4.1) whose importance derives from the fact (proven in proposition 4.3) that any purely temporal transformation of a texture quilt is microbalanced. Concrete methods are provided for constructing *binary* and *sinusoidal* texture quilts that display consistent motion.

In Section 5, these construction methods are applied in an experiment designed to demonstrate the effectiveness of three textural properties as carriers of motion information. The textural properties are (i) spatial frequency variation, (ii) orientation variation, and (iii) variation between perceptually distinct textures with identical expected energy spectra.

2. Preliminaries

This section states the background facts presupposed by the main discussion of the paper.

2.1. *Discrete Dynamic Visual Stimuli*

Notation. Let $\mathbb{R}$ denote the real numbers, and $\mathbf{Z}$ ($\mathbf{Z}^+$) the integers (positive integers). We use square brackets to enclose arguments of discrete functions, and parentheses to enclose arguments of continuous functions.

The Range of a Stimulus. We want the term "stimulus" to refer not only to the luminance function submitted as input to the retina, but to any physiologically reasonable transformation of the spatiotemporal luminance function which might be submitted as input to a component processor of the visual system. Consequently, although luminance is physically a nonnegative quantity, we do not apply this constraint to the class of functions we admit as stimuli. We allow stimuli to take values throughout the positive and negative real numbers.

The Domain of a Stimulus. To remain close to our intuitions about neurally realized visual processors, we take stimuli to be a functions of the discrete domain $\mathbf{Z}^3$ (where the dimensions correspond to horizontal and vertical space, and time). In addition, for mathematical convenience, and without loss of physiological plausibility, we require a stimulus to be 0 almost everywhere in its (infinite) domain.

The Definition of a Stimulus. We call any function $I: \mathbf{Z}^3 \to \mathbb{R}$ a *stimulus* provided $I[x, y, t] = 0$ for all but finitely many points of $\mathbf{Z}^3$.

We shall be considering stimuli as functions of two spatial dimensions x, y and time t.

Stimulus Contrast. As is now well established (e.g., Shapley & Enroth-Cugell, 1984), early retinal gain-control mechanisms pass not stimulus luminance, but rather a signal approximating stimulus *contrast*, the normalized deviation at each time t of luminance at each point (x, y) in the visual field from a "background level," or "level of adaptation," which reflects the average luminance over points proximal to (x, y, t) in space and time. Because the transformation from luminance to contrast is a processing stage that is general to all of vision, we shall drop reference to mean luminance L_0, and characterize L only by its *contrast modulation function*, C:

$$C = \frac{L}{L_0} - 1. \tag{1}$$

What we argue in this paper is that the broad-band spatial filtering that mediates the step from luminance to contrast is succeeded by additional filtering stages in which a number of *narrowly tuned* spatial filters are applied to the visual signal, their output rectified, and the resulting spatiotemporal signal processed for motion information.

The History of a Stimulus at a Point in Space. For any stimulus I, any point $(x, y) \in \mathbf{Z}^2$, we define $I_{(x,y)}$, the *history of I at* (x, y), by setting

$$I_{(x,y)}[t] = I[x, y, t] \tag{2}$$

for all $t \in \mathbf{Z}$.

Space-Time Separable Stimuli. A stimulus I is called *space-time* separable iff I can be expressed as the product of a spatial function $f: \mathbf{Z}^2 \to \mathbf{R}$ and a temporal function $g: \mathbf{Z} \to \mathbf{R}$: For all $(x, y, t) \in \mathbf{Z}^3$, $I[x, y, t] = f[x, y]\, g[t]$.

The Fourier Transform of a Stimulus. Because any stimulus I is nonzero at only a finite number of points, the energy in I is finite, implying that I has a well-defined Fourier transform.

We denote I's Fourier transform by $\hat{I}$: writing i for the complex number $(0, 1)$,

$$\hat{I}(\omega, \theta, \tau) = \sum_{x} \sum_{y} \sum_{t} I[x, y, t]\, e^{-i(\omega x + \theta y + \tau t)}. \tag{3}$$

Although $\hat{I}$ is defined for all real numbers ω, θ, τ, it is periodic over 2π in each argument. This fact is reflected in the inverse transform:

$$I[x, y, t] = \frac{1}{(2\pi)^3} \int_{-\pi}^{+\pi} \int_{-\pi}^{+\pi} \int_{-\pi}^{+\pi} \hat{I}(\omega, \theta, \tau)\, e^{i(\omega x + \theta y + \tau t)}\, d\omega\, d\theta\, d\tau. \tag{4}$$

In the Fourier domain, we consistently use ω to index frequencies relative to x, θ frequencies relative to y, and τ frequencies relative to t.

The Function **0**. We write **0** for any function that assigns 0 to each element in its domain. Thus, **0** defined on $\mathbf{Z}^3$ is the stimulus that is zero throughout space and time. We also write **0** for the temporal function that sets $\mathbf{0}[t] = 0$ for all $t \in \mathbf{Z}$.

2.2. *Mappings and Stimulus Transformations*

Let Ω be the set of all real-valued functions of $\mathbf{Z}^3$, and call any function of Ω into Ω a *mapping*. (We shall need the general notion of a mapping only briefly in order to specify the subset of well-behaved mappings called transformations.) For any mapping M and any $I \in \Omega$, $M(I)$ is a real-valued function of $\mathbf{Z}^3$; accordingly, we write $M(I)[x, y, t]$ for the value of $M(I)$ at any point $(x, y, t) \in \mathbf{Z}^3$.

If it is continuous, a function $f: \mathbb{R} \to \mathbb{R}$ submits to a wide range of useful operations. For instance, if f is continuous, it can be integrated over any finite interval. Of course, f need not be continuous to meet this condition. For instance, f is integrable over any finite interval if f is discontinuous at only a finite number of points in any finite interval. If f is integrable over any finite interval, and if f also is bounded, then for any function g for which $\int_{\mathbb{R}} g$ converges, $\int_{\mathbb{R}} fg$ also converges. In particular, $\int_{\mathbb{R}} fg$ converges if g is a density function. For the results reported here, we restrict our attention to a special class of mappings, which we shall call stimulus transformations, that have properties analogous to those of the well-behaved function f. We specify these desirable properties in the following paragraph.

Continuous Mappings; Finitely Integrable Mappings; Bounded Mappings. For any $I \in \Omega$, any $p \in \mathbb{R}$, any $\psi \in \mathbf{Z}^3$, we write $I_{\psi \leftarrow p}$ for the element of Ω that is identical to I at all locations of $\mathbf{Z}^3$ except ψ, where it takes the value p. Any mapping M is called *continuous* if $M(I_{\psi \leftarrow p})[\zeta]$ is a continuous function of p for any $I \in \Omega$, and any $\psi, \zeta \in \mathbf{Z}^3$. M is called *finitely integrable* if, for any such I, ψ, and ζ, $M(I_{\psi \leftarrow p})[\zeta]$ is an integrable function of p over any finite interval. Finally, M is called *bounded* if, for any such I, ψ, and ζ, $M(I_{\psi \leftarrow p})[\zeta]$ is a bounded function of p over the set of real numbers.

DEFINITION OF A STIMULUS TRANSFORMATION. A *stimulus transformation* (which we shall often refer to simply as a *transformation*) is a bounded, finitely integrable, mapping T such that $T(S)$ is a stimulus for any stimulus S, and $T(\mathbf{0}) = \mathbf{0}$.

There are other reasonable constraints we might impose on the notion of a stimulus transformation. For instance, we might require a stimulus transformation to be time-invariant and causal. However, we do not include these conditions in our definition because they are not required for the results we report.

Purely Temporal Stimulus Transformations. Let Ω_t be the set of all functions mapping $\mathbf{Z}$ into $\mathbb{R}$. A transformation H is called *purely temporal* iff there exists a function $H_t: \Omega_t \to \Omega_t$ such that for any stimulus I, any $(x, y, t) \in \mathbf{Z}^3$,

$$H(I)[x, y, t] = H_t(I_{(x,y,\cdot)})[t]. \tag{5}$$

That is, the value at the point $(x, y, t) \in \mathbf{Z}^3$ that results from applying H to I depends only on the history of I at (x, y). Since it is obvious from the context, we drop the distinction between H and H_I, and allow H to be applied both to full-fledged stimuli and to simple functions of time. Thus, for any temporal function $P: \mathbf{Z} \to \mathbb{R}$, we shall write $H(P)$ to indicate the temporal function $H_I(P)$.

We shall be particularly concerned with two types of transformations: *pointwise* transformations and *linear, shift-invariant* transformations.

Pointwise Transformations and Rectifiers. For any functions $f: A \to B$ and $g: B \to C$, the *composition* $g \bullet f: A \to C$ is given by

$$g \bullet f(a) = g(f(a)) \tag{6}$$

for any $a \in A$. For any $f: \mathbb{R} \to \mathbb{R}$, we call the mapping $f\bullet$, yielding the spatiotemporal function $f \bullet I$ when applied to stimulus I, a *pointwise* mapping (because its output value at any point in space-time depends only on its input value at that point).

As is evident, $f\bullet$ is a transformation iff (i) $f(0) = 0$, (ii) f is bounded on $\mathbb{R}$, and (iii) f is integrable over any bounded real interval. A transformation $f\bullet$ is called a *positive half-wave rectifier* if f is monotonically increasing, and $f[v] = 0$ for all $v \leqslant 0$; $f\bullet$ is called a *negative half-wave rectifier* if f is monotonically decreasing, and $f[v] = 0$ for $v \geqslant 0$. Finally, $f\bullet$ is called a *full-wave rectifier* if f is a monotonically increasing function of absolute value.

Linear, Shift-Invariant (LSI) Transformations. For any offset $\psi \in \mathbf{Z}^3$, define the mapping S^ψ by

$$S^\psi(I)[\zeta] = I[\zeta - \psi] \tag{7}$$

for any $I \in \Omega$. Thus $S^\psi(I)$ is derived by shifting I by the offset ψ in $\mathbf{Z}^3$. Any mapping M is called *shift-invariant* iff

$$S^\psi(M(I)) = M(S^\psi(I)) \tag{8}$$

for any $\psi \in \mathbf{Z}^3$, any $I \in \Omega$. In addition, M is *linear* iff for any $I, J \in \Omega$, any real numbers κ and λ

$$M(\kappa I + \lambda J) = \kappa M(I) + \lambda M(J). \tag{9}$$

As is well known, any linear, shift-invariant (LSI) transformation can be expressed as a *convolution*, which is defined for any $u \in \mathbf{Z}^3$ by

$$(k * I)[u] = \sum_{\mathbf{Z}^3} k[u - v]\, I[v], \tag{10}$$

for some $k: \mathbf{Z}^3 \to \mathbb{R}$. The function k is called the *impulse response* of the transformation $k*$.

2.3. *Random Stimuli*

For any real random variable X with density f, we write $E[X]$ for the *expectation of* X:

$$E[X] = \int_{-\infty}^{\infty} x f(x)\, dx. \tag{11}$$

The notion of a random stimulus generalizes that of a (nonrandom) stimulus in that the values assigned points in space-time by a random stimulus are random variables (with finite variances) rather than constants.

DEFINITION OF A RANDOM STIMULUS. Call any family $\{R[x, y, t] \mid (x, y, t) \in \mathbf{Z}^3\}$ of jointly distributed random variables a *random stimulus* provided

(i) $R[x, y, t]$ is constant and equal to 0 for all but finitely many $(x, y, t) \in \mathbf{Z}^3$, and

(ii) $E[R[x, y, t]^2]$ exists for all $(x, y, t) \in \mathbf{Z}^3$.

As with nonrandom stimuli, we write $\bar{R}$ for the Fourier transform of any random stimulus R; and, for any $\chi = (x, y) \in \mathbf{Z}^2$ we write R_χ for the temporal random function defined by

$$R_\chi[t] = R[\chi, t] \tag{12}$$

for all times $t \in \mathbf{Z}$.

Space-Time Separable Random Stimuli. We call a random stimulus R *space-time separable* iff R is space-time separable with probability 1.

Constant Stimuli. Any ordinary stimulus can be regarded as a random stimulus that does not vary across independent realizations. We call such unvarying stimuli *constant*.

The Motion-from-Fourier-Components Principle. *Parseval's relation* states that the energy in a stimulus is proportional to the energy in its Fourier transform. Individual spatiotemporal Fourier components are drifting sinusoidal gratings. Thus, we can add up the energy in a dynamic visual stimulus either point-by-point in space-time, or drifting sinusoid by drifting sinusoid. A commonly encountered rule of thumb (van Santen & Sperling, 1985; Watson & Ahumada, 1983b; Watson, Ahumada, & Farrell, 1986) for predicting the apparent motion of an arbitrary stimulus $I[x, y, t] = I[x, t]$ (constant in the vertical dimension of space), is the *motion-from-Fourier-components* principle: For I regarded as a linear combination of drifting sinusoidal gratings, if most of I's energy is contributed by rightward-drifting gratings, then perceived motion should be to the right. If most of the energy resides in the leftward-drifting gratings, perceived motion should be to the left. Otherwise I should manifest no decisive motion in either direction.

Drift-Balanced Random Stimuli. The class of *drift-balanced* random stimuli (Chubb & Sperling, 1987, 1988) provides a rich pool of counterexamples to the motion-from-Fourier-components principle. A random stimulus R is drift balanced iff the expected energy in R of each drifting sinusoidal component is equal to the expected energy of the component of the same spatial frequency, drifting at the same rate, but in the opposite direction. The term *drift balanced* is defined formally as follows.

DEFINITION OF A DRIFT-BALANCED RANDOM STIMULUS. Call any random stimulus R *drift balanced* iff

$$E[|\bar{R}(\omega, \theta, \tau)|^2] = E[|\bar{R}(\omega, \theta, -\tau)|^2] \tag{13}$$

for all $(\omega, \theta, \tau) \in \mathbb{R}^3$.[1]

Thus, for any class of spatiotemporal linear receptors tuned to stimulus energy in a certain spatiotemporal frequency band, a drift-balanced random stimulus will, on the average, stimulate equally well those receptors tuned to the corresponding band of opposite temporal orientation.

Microbalanced Random Stimuli. Consider the following two-flash stimulus S: In flash 1, a bright spot (call it Spot 1) appears. In flash 2, Spot 1 disappears, and two new spots appear, one to the left and one symmetrically to the right of Spot 1. As one might suppose, S is drift balanced. On the other hand, it is equally clear that a Fourier motion detector whose spatial reach encompasses the location of Spot 1 and only *one* of the Spots in flash 2 may well be stimulated in a fixed direction by S. Thus, although S is drift balanced, some Fourier motion detectors may be stimulated strongly and systematically by S. These detectors can be differentially selected by *spatial windowing*, and thereby the drift-balanced stimulus S is converted into a non-drift-balanced stimulus by multiplying it by an appropriate space-time separable function. The following subclass of drift-balanced random stimuli cannot be made non-drift-balanced by space-time separable windowing.

DEFINITION OF A MICROBALANCED RANDOM STIMULUS. Call any random stimulus I *microbalanced* iff the product WI is drift balanced for any space-time separable function W.

One can think of the multiplying function W as a "window" through which a spatiotemporal subregion of I can be "viewed" in isolation. The space-time separability of W ensures that W is "transparent" with respect to the motion-content of the region to which it is applied: W does not distort I's motion with any motion content of its own. The fact that I is microbalanced means that any subregion of I encountered through a "motion-transparent window" is drift balanced.

[1] For a proof that the expected energy of the Fourier transform of any random stimulus is everywhere well defined see Chubb & Sperling (1988, Appendix A).

The following characterization of the class of microbalanced random stimuli, and all other results stated without proof in this section, are from Chubb and Sperling (1988).

2.3.1. *A random stimulus I is microbalanced if and only if*

$$E[I[x, y, t]\, I[x', y', t'] - I[x, y, t']\, I[x', y', t]] = 0 \tag{14}$$

for all $x, y, t, x', y', t' \in \mathbf{Z}$.

Some other relevant facts about microbalanced random stimuli:

2.3.2. *For any independent microbalanced random stimuli I and J,*

I. *the product IJ is microbalanced,*

and

II. *the convolution $I * J$ is microbalanced.*

2.3.3. (a) *Any space-time separable random stimulus is microbalanced;* (b) *any constant microbalanced stimulus is space-time separable.*

The following result is useful in constructing a wide range of microbalanced random stimuli which display striking apparent motion.

2.3.4. *Let Γ be a family of pairwise independent, microbalanced random stimuli, all but at most one of which have expectation* **0**. *Then any linear combination of Γ is microbalanced.*

Reichardt Detectors and Microbalanced Random Stimuli. Two Fourier motion detectors proposed for psychophysical data (Adelson & Bergen, 1985; Watson & Ahumada, 1983a, 1983b) can be recast as *Reichardt detectors* (Adelson & Bergen, 1985; van Santen & Sperling, 1985). The Reichardt detector has many useful properties as a motion detector without regard to its specific instantiation (van Santen & Sperling, 1984, 1985).

Figure 1 shows a diagram of the Reichardt detector. It consists of spatial receptors characterized by spatial functions f_1 and f_2, temporal filters g_1* and g_2*, multipliers, a differencer, and another temporal filter $h*$. The spatial receptors f_i, $i = 1, 2$, act on the input stimulus I to produce intermediate outputs,

$$y_i[t] = \sum_{x, y \in \mathbf{Z}} f_i[x, y]\, I[x, y, t]. \tag{15}$$

At the next stage, each temporal filter g_j* transforms its input y_i $(i, j = 1, 2)$, yielding four temporal output functions: $g_j * y_i$. The left and right multipliers then compute the products

$$[y_1 * g_1[t]][y_2 * g_2[t]] \quad \text{and} \quad [y_1 * g_2[t]][y_2 * g_1[t]]. \tag{16}$$

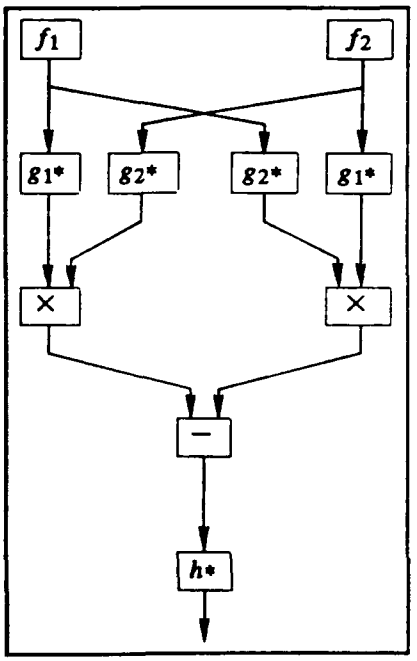


FIG. 1. The Reichardt detector. Let I be a random stimulus. Then, in response to I, for $i = 1, 2$, the box containing the spatial function $f_i: \mathbf{Z}^2 \to \Re$, outputs the temporal function, $\sum_{x,y \in \mathbf{Z}} f_i[x, y] I[x, y, t]$; each of the boxes marked g_i* outputs the convolution of its input with the temporal function $g_i: \mathbf{Z} \to \Re$; each of the boxes marked with a multiplication sign outputs the product of its inputs; the box marked with a minus sign outputs its left input minus its right; and the box containing $h*$ outputs the convolution of its input with the temporal function $h: \mathbf{Z} \to \Re$. To see how the Reichardt detector senses motion, suppose f_2 is identical to f_1, but shifted in space by some offset, and suppose the filters g_1* do not alter their input, while the filters g_2* simply delay their input by some amount δ_t of time. Then a rigidly translating pattern moving in the direction of box f_2's offset from box f_1 will elicit some time-varying response from box f_1, and the same response a short time later from box f_2. If that "short time later" is precisely δ_t, the output of the righthand multiplier will be positive as long as the pattern keeps drifting. This will result in a net negative Reichardt detector output. If the pattern drift is in the opposite direction, the detector response will be positive.

respectively, and the differencer subtracts the output from the right multiplier from that of the left multiplier:

$$D[t] = [y_1 * g_1[t]][y_2 * g_2[t]] - [y_1 * g_2[t]][y_2 * g_1[t]]. \tag{17}$$

The final output is produced by applying the filter $h*$, whose purpose is to smooth the time-varying, differencer output D. Since many Fourier mechanisms can be expressed as, or closely approximated by, Reichardt detectors (Adelson & Bergen, 1985, 1986; van Santen & Sperling, 1985), the following characterization of the class of microbalanced stimuli can be regarded as the cornerstone of the claim that microbalanced random stimuli bypass Fourier motion mechanisms.

2.3.5. *For any random stimulus I, the following conditions are equivalent:*

(I) *I is microbalanced.*

(II) *The expected response of every Reichardt detector to I is 0 at every instant in time.*

Proof. Chubb & Sperling (1988) proved that I implies II. To obtain the reverse implication, note that if II holds, then, in particular, for any points (x, y), $(x', y') \in \mathbf{Z}^2$ and any $\delta_t \in \mathbf{Z}$, the expected response to I is the temporal function $\mathbf{0}$ for a particular simple Reichardt detector that computes

$$I[x, y, t]\, I[x', y', t - \delta_t] - I[x, y, t - \delta_t]\, I[x', y', t]. \tag{18}$$

This Reichardt detector is constructed by making (i) f_1 (of Fig. 1) the function that takes the value 1 at (x, y) and 0 everywhere else, (ii) f_2 the function that takes the value 1 at (x', y') and 0 everywhere else, (iii) each of g_1* and $h*$ the identity transformation, and (iv) g_2* the filter that delays its input by δ_t units of time. However, if the expected response to I is 0 throughout time for any such Reichardt detector, then Eq. (14) holds, and proposition 2.3.1 implies that I is microbalanced. ∎

3. Random Stimuli Microbalanced under All Pointwise Transformations

The main purpose of this paper is to provide tools for differentially stimulating specific types of non-Fourier motion mechanisms without engaging either Fourier mechanisms or other types of non-Fourier mechanisms. A non-Fourier motion mechanism is one that applies an initial nonlinear transformation to the visual signal and subjects the output to standard motion analysis. In this section, we provide some results relevant to the psychophysical problem of stimulating non-Fourier mechanisms whose initial transformation is nonpointwise without engaging any mechanism whose initial transformation is pointwise. The main finding is stated in proposition 3.2, which provides necessary and sufficient conditions for a random stimulus I to be such that $t \bullet I$ is microbalanced for any pointwise transformation $t\bullet$. In Section 4 we apply this result to construct random stimuli (texture quilts) which are microbalanced, and are, moreover, guaranteed to remain microbalanced after any purely temporal transformation. Such stimuli are useful for selectively stimulating non-Fourier motion mechanisms that extract motion information from stimuli that have undergone nonlinear *spatial* stimulus transformations.

We begin by considering an example of a stimulus (Chubb & Sperling, 1987, 1988) that is microbalanced under all pointwise transformations, but whose motion can be revealed by a purely temporal nonlinear transformation.

3.1. *Stimulus J: Traveling Reversal of a Random Black-or-White Vertical Bar Pattern.* Let $M \in \mathbf{Z}^+$. We construct the random stimulus J of $M + 1$ frames

indexed 0, 1, ..., M, each of which contains M vertical bars, indexed 1, 2, ..., M from left to right. In frame 0 of stimulus J, all M vertical bars first appear. The contrast of each bar is 1 or -1 with equal probability, and bar contrasts are jointly independent. In each successive frame m, $m = 1, 2, \ldots, M$, the mth rectangle flips its contrast to 1 if its previous contrast was -1; otherwise it flips from 1 to -1. In frame 1, rectangle 1 flips contrast; in frame 2, rectangle 2 flips, and in successive frames, successive rectangles flip contrast from left to right, until the Mth rectangle flips in frame M, after which all the rectangles turn off. An xt cross-section of frames 0 to M of J is shown in Fig. 2a.

The traveling contrast-reversal, stimulus J, is easily expressed as a sum of pairwise independent, space-time separable random stimuli, all with expectation **0**; thus propositions 2.3.3a and 2.3.4 imply that J is microbalanced. Moreover, it is easy to see that, because J's frames are comprised of only two values, any pointwise transformation of J merely serves to rescale each of J's frames, and to shift it by a constant; that is, for any $f: \mathbb{R} \to \mathbb{R}$, $f \bullet J = \lambda J + K$, where $\lambda \in \mathbb{R}$, and K is a stimulus

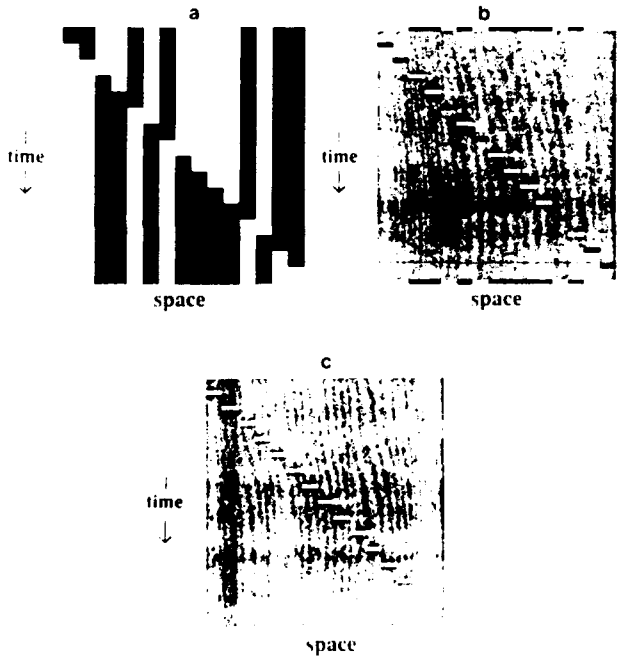


FIG. 2. Exposing the motion of the traveling contrast-reversal of the random black-or-white vertical bar pattern J to standard motion-analysis. (a) An xt cross-section of J. (b) An xt cross-section of the partial derivative of J with respect to time. (c) An xt cross-section of $\partial J/\partial t$. Each of J and $\partial J/\partial t$ is microbalanced. However, $\partial J/\partial t$ is not. In particular, $\partial J/\partial t$ has most of its energy at those frequencies whose velocity is equal to the velocity of the traveling contrast-reversal.

that assigns a constant value across all points at which J is nonzero. Clearly, $f \bullet J$ is another microbalanced random function (this follows easily from proposition 2.3.4). Thus, pointwise transformations fail to expose J's motion.

Exposing J's Motion to Standard Analysis. Perhaps the simplest way to extract J's motion is to full-wave rectify the partial derivative of J taken with respect to time. The stages of this transformation are illustrated in Figs. 2b and 2c. Figure 2b shows $\partial J / \partial t$. This function is itself microbalanced (propositions 2.3.2II and 2.3.3a imply that any purely temporal LSI transformation of a microbalanced random stimulus is microbalanced). However, $\partial J / \partial t$ (Fig. 2c) has most of its energy at those spatiotemporal frequencies whose velocity is equal to the velocity of the traveling contrast-reversal whose motion we wish to detect. Thus we see that, although J's motion cannot be exposed to standard analysis by a simple pointwise transformation, a temporal linear filter followed by a pointwise nonlinearity does suffice.

We turn now to the problem of stipulating the general conditions that a random stimulus I must satisfy so that $f \bullet I$ will be microbalanced for any pointwise transformation $f\bullet$. Call any random stimulus I *microbalanced under* a given transformation T iff $T(I)$ is microbalanced.

We state the following basic proposition (3.2) and its subsequent corollary (3.3) for continuously distributed random stimuli. The corresponding result for discretely distributed random stimuli is simpler and should be evident.

3.2. NECESSARY AND SUFFICIENT CONDITIONS FOR A RANDOM STIMULUS TO BE MICROBALANCED UNDER ALL POINTWISE TRANSFORMATIONS. *Let I be a random stimulus such that for any (x, y, t), $(x', y', t') \in \mathbf{Z}^3$, $(I[x, y, t], I[x', y', t'])$ has a continuous joint density. Then the following conditions are equivalent:*

(1) *I is microbalanced under all pointwise transformations.*

(2) *For all $x, y, t, x', y', t' \in \mathbf{Z}$, the joint density f of $(I[x, y, t], I[x', y', t'])$ and the joint density g of $(I[x, y, t'], I[x', y', t])$ satisfy*

$$f(p, q) + f(q, p) = g(p, q) + g(q, p) \tag{19}$$

for any $p, q \in \mathbb{R}$ such that $p \neq 0$ and $q \neq 0$.

Proof. Set $\kappa = I[x, y, t]$, $\lambda = I[x', y', t']$, $\gamma = I[x, y, t']$, and $\nu = I[x', y', t]$. Thus, (κ, λ) is distributed in $\mathbb{R}^2$ with density f and (γ, ν) is distributed with density g.

(2) implies (1): By definition of any pointwise transformation $h\bullet$, we have $h(0) = 0$. Thus we need integrate only over values of κ and λ which are both nonzero in computing the expectation $E[h(\kappa)\, h(\lambda)]$. In particular, if Eq. (19) is satisfied for all $p \neq 0$ and $q \neq 0$, then $h \bullet I$ is microbalanced since

$$E[h(\kappa)\,h(\lambda)] = \frac{1}{2}\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(p)\,h(q)\,f(p,q)\,dp\,dq + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(q)\,h(p)\,f(q,p)\,dq\,dp\right]$$

$$= \frac{1}{2}\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(p)\,h(q)\,f(p,q)\,dp\,dq + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(p)\,h(q)\,f(q,p)\,dp\,dq\right]$$

$$= \frac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(p)\,h(q)(f(p,q)+f(q,p))\,dp\,dq$$

$$= \frac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(p)\,h(q)(g(p,q)+g(q,p))\,dp\,dq = E[h(\gamma)\,h(\nu)]. \quad (20)$$

(Note: the boundedness and finite integrability of $h\bullet$ ensure that these expectations exist.)

(Not (2) implies not (1)): On the other hand, suppose Eq. (19) fails for some $x, y, t, x', y', t' \in \mathbf{Z}$. One way in which this might happen is if $f(r,r) > g(r,r)$ for some nonzero $r \in \mathbb{R}$. In this case, there exists a neighborhood N of r, not including 0, such that $f(m,n) > g(m,n)$ for all $m, n \in N$. Thus, for the function $h: \mathbb{R} \to \mathbb{R}$ defined by

$$h(n) = \begin{cases} 1 & \text{if } n \in N, \\ 0 & \text{otherwise,} \end{cases} \quad (21)$$

$h\bullet$ is a pointwise transformation (the function h is bounded on $\mathbb{R}$, finitely integrable, and $h(0) = 0$). However, $h\bullet I$ is not microbalanced since

$$E[h(\kappa)\,h(\lambda)] = \int_{N}\int_{N} f(m,n)\,dm\,dn > \int_{N}\int_{N} g(m,n)\,dm\,dn$$

$$= E[h(\gamma)\,h(\nu)]. \quad (22)$$

To recapitulate, if Condition 2 fails because there exists a nonzero $r \in \mathbb{R}$ for which $f(r,r) \neq g(r,r)$, then Condition 1 fails (I is not microbalanced under all pointwise transformations).

The only other way in which Condition 2 can fail is if $f(r,r) = g(r,r)$ for all $r \neq 0$ in $\mathbb{R}$, but for some $p, q \in \mathbb{R}$, with neither p nor q equal to 0, $f(p,q) + f(q,p) > g(p,q) + g(q,p)$. In this case, we obtain disjoint neighborhoods M of p and N of q, neither including 0, such that

$$f(m,n) + f(n,m) > g(m,n) + g(n,m) \quad (23)$$

A random stimulus microbalanced under all pointwise transformations, but quite different from J of example 3.1 is the following, suggested by J. Lappin (1989).

3.4. *Stimulus K: Rotating Random-Dot Cylinder.* Construct K by taking the parallel projection of a set of points on (and or inside) the surface of a cylinder rotating around a vertical axis. Let the contrast values of the points be independent, identically distributed random variables. As is well known, when properly constructed, K can display a very strong kinetic depth effect, with dots moving in one direction seen as being in the front of the axis of rotation, and dots moving in the other direction seen as being in the back (Dosher, Landy, & Sperling, 1989; Ullman, 1979). Nonetheless, K is microbalanced under all pointwise transformations: All of K's systematic motion is horizontal; thus, we can drop reference to y, and note that for any x, t, x', t', the joint distribution of $(K[x, t], K[x', t'])$ is identical to that of $(K[x, t'], K[x', t])$. Hence, by Corollary 3.3, Condition 3, K is microbalanced under all pointwise transformations.

4. Texture Quilts

The rest of this paper is devoted to illustrating how the results of Section 3 can be applied to construct stimuli which display consistent apparent motion that cannot be exposed to standard analysis by any purely temporal transformation. Specifically, we demonstrate several motion-displaying stimuli, called *texture quilts* (Definition 4.1), that are microbalanced under all purely temporal transformations.

As illustrated in Fig. 3, the simplest transformations that suffice to expose the motion of texture quilts to standard analysis involve a purely spatial linear filter $s*$ followed by a rectifier $r\bullet$:

$$T(Q) = r\bullet(s * Q). \tag{31}$$

The spatial filter $s*$ will respond with varying energy throughout regions of the visual field, depending on whether or not the textures to which it is tuned populate those regions. However, the output of a linear filter to a texture is positive or negative depending on the local phase of the texture. The purpose of rectification is to transform regions of high-variance $s*$ response into regions of high average value, thus ensuring that the rectified output registers the presence or absence of texture, independent of phase. The result $T(Q)$ is a spatiotemporal function whose value reflects the local texture preferences of $s*$ in the visual field as a function of time (Bergen & Adelson, 1988; Caelli, 1985).[2]

[2] In general, a spatial linear filter followed by a pointwise nonlinearity can have arbitrarily high order Volterra kernels, depending on the order of the Taylor series of the pointwise transformation. However, if we take the rectifier of step (2) to be $\mathrm{Rect}(x) = x^2$, then this squared output of a spatial filter is a second order spatial transformation. Standard motion analysis is yet another second order transformation. Thus, when we subject the squared filter output to standard motion analysis, we are applying a fourth order operator.

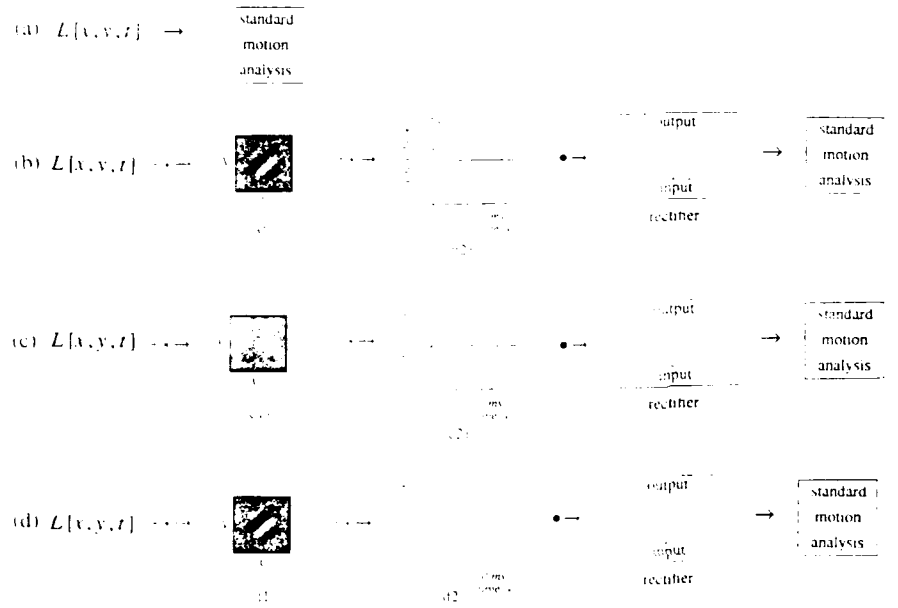


FIG. 3. Fourier and non-Fourier motion mechanisms. (a) Fourier motion mechanisms apply standard motion-analysis directly to the luminance signal L. (b), (c), and (d) Non-Fourier mechanisms apply standard motion-analysis to a nonlinear transformation of luminance. (b) A simple non-Fourier mechanism applies a signal transformation comprised of a spatiotemporal linear filter, followed by a pointwise nonlinearity. The *'s indicate spatial and temporal convolution, respectively, and • indicates function composition. The filtering performed in (b) is roughly pointwise in time (the temporal impulse response b2 approximates an impulse), and the nonlinearity applied is a full-wave rectifier. This system (with appropriately chosen spatial filter, b1) will extract the motion of the texture quilts shown in Figs. 4b, 5d, 6c, and 6d. It will not extract the motion of stimulus J, the traveling contrast-reversal of the random vertical bar pattern shown in Fig. 2a. (c) A spatially pointwise (the spatial impulse response c1 approximates an impulse), system with a flicker-sensitive temporal filter and a full-wave rectifier. Because of the flicker sensitivity, this mechanism will extract the motion of the traveling contrast-reversal of the random vertical bar pattern shown in Fig. 2a but not the motion of the texture quilts shown in Figs. 4b, 5d, 6c, and 6d. (d) The temporal filter d2 averages the temporal filters b2 and c2, and the pointwise nonlinearity is a full-wave rectifier. With an appropriate spatial filter d1, the non-Fourier system extracts the motion of any corresponding texture quilt as well as the motion of the traveling contrast-reversal of the random vertical bar pattern shown in Fig. 2a. However, it would be less well suited to these tasks than the detectors shown in (b) and (c) whose temporal filters it averages.

The essential trick in all the quilt examples we consider is to patch together various brief displays of static, random texture, taking appropriate measures to ensure that the resultant stimulus satisfies the following definition.

4.1. DEFINITION OF A TEXTURE QUILT. Let $A = \mathbf{Z}^2$ be a set of points in space, and let $t_0, t_1, \ldots, t_N$ be a strictly increasing sequence of times, with $T = \{t : t_0 \leq t < t_N\}$. Call any random stimulus Q satisfying the following conditions a *texture quilt*:

(i) Q assigns 0 to all points outside $A \times T$.

(ii) For $i = 0, 1, \ldots, N-1$, the random values assigned by Q to points in A at time t_i remain unchanged until time t_{i+1}.

(iii) *Independence.* For $i = 0, 1, \ldots, N-1$, the random substimuli Q^i, defined, for all points α in space and all times t, by

$$Q^i[\alpha, t] = \begin{cases} Q[\alpha, t] & t_i \leqslant t < t_{i+1},\ \alpha \in A \\ 0 & \text{otherwise.} \end{cases} \tag{32}$$

are jointly independent.

(iv) *Symmetry.* For any $\alpha, \beta \in A$, and any $t \in T$, the joint distribution of $(Q[\alpha, t], Q[\beta, t])$ is identical to the joint distribution of $(Q[\beta, t], Q[\alpha, t])$.

Terminology. Call A and T respectively Q's *spatial* and *temporal regions of activity*, and for $i = 0, 1, \ldots, N-1$, call $\{t \mid t_i \leqslant t < t_{i+1}\}$ the ith *timeblock* of Q.

The empirical usefulness of texture quilts derives from proposition 4.3 in conjunction with the fact that it is easy to construct various sorts of texture quilts which display consistent apparent motion across independent realizations. The proof of proposition 4.3 is eased by the following

4.2. LEMMA. *Let Q be a texture quilt with spatial region of activity A. Then for any $\alpha, \beta \in A$, the pair of temporal functions (Q_α, Q_β) is distributed identically to the reverse pair (Q_β, Q_α)*

Proof. From Definition 4.1(i) and (ii), note that for temporal functions P and R, the density of the joint assignment $(Q_\alpha, Q_\beta) = (P, R)$ is 0 unless each of P and R is constant throughout each time block, and 0 outside T. Thus, any P and R for which the joint assignment $(Q_\alpha, Q_\beta) = (P, R)$ has nonzero density are completely determined by the values $P[t_i] = p_i$, and $R[t_i] = r_i$, for $i = 0, 1, \ldots, N-1$; for f_i the joint density of $(Q_\alpha[t_i], Q_\beta[t_i])$. Definition 4.1(iii) thus implies that the density of the joint assignment $(Q_\alpha, Q_\beta) = (P, R)$ is

$$\prod_{i=0}^{N-1} f_i(p_i, r_i). \tag{33}$$

But by Definition 4.1(iv), the quantity (33) is equal to

$$\prod_{i=0}^{N-1} f_i(r_i, p_i), \tag{34}$$

which is the density of the reverse occurrence that $(Q_\beta, Q_\alpha) = (P, R)$. ∎

4.3. TEXTURE QUILTS ARE MICROBALANCED UNDER PURELY TEMPORAL TRANSFORMATIONS. I. *Any texture quilt with a continuous joint density is microbalanced under all purely temporal, continuous transformations.*

II. *Any discretely distributed texture quilt is microbalanced under all purely temporal transformations.*

Proof of I. Let Q be a texture quilt with a continuous joint density, and let Φ be an arbitrary purely temporal, continuous transformation. We must prove that $\Phi(Q)$ is microbalanced. We can, of course, accomplish this by proving that $\Phi(Q)$ is microbalanced under all pointwise transformations (since, in particular, the identity transformation is pointwise). This turns out to be a convenient approach.

Let α, β be points in space, and let t and u be points in time. Because Φ is bounded and continuous and Q has a continuous joint density, we know that the joint density f of $(\Phi(Q)[\alpha, t], \Phi(Q)[\beta, u])$ and the joint density g of $(\Phi(Q)[\beta, t], \Phi(Q)[\alpha, u])$ both exist and are continuous on $\mathbb{R}^2$. We shall show for any $(p, r) \in \mathbb{R}^2$ with neither p nor r equal to 0, that either $f(p, r) = g(p, r)$ or $f(p, r) = g(r, p)$. The proposition will then follow from Corollary 3.3.

Case 1. At least one of α or β is outside A. Suppose α is outside A. Then by Definition 4.1(i), $Q_\alpha = \mathbf{0}$; hence $\Phi(Q)[\alpha, t] = \Phi(Q)[\alpha, u] = 0$. Consequently, $f(p, r) = g(r, p) = 0$ whenever $p \neq 0$. Thus Eq. (29) holds vacuously, with

$$f(p, r) = g(r, p) = 0 \qquad \text{for all } p, r \in \mathbb{R},\ p \neq 0,\ r \neq 0. \tag{35}$$

Case 2. Both α and β are in A. Let F be the joint density of (Q_α, Q_β) and G be the joint density of (Q_β, Q_α). By Lemma 4.2, $F = G$. Clearly, then, for F_Φ the joint density of $(\Phi(Q_\alpha), \Phi(Q_\beta))$ and G_Φ the joint density of $(\Phi(Q_\beta), \Phi(Q_\alpha))$, it follows that $F_\Phi = G_\Phi$. For any $p, r \in \mathbb{R}$, recall that $f(p, r)$ is the density of the co-occurrence that $\Phi(Q)[\alpha, t] = p$, and $\Phi(Q)[\beta, u] = r$, but this is precisely the density of the event that $(\Phi(Q_\alpha)[t], \Phi(Q_\beta)[u]) = (p, r)$. This density, however, is equal to the integral of F_Φ over all pairs of temporal functions (P, R) such that $P[t] = p$ and $R[u] = r$. Similarly, $g(p, r)$ is the density of the co-occurrence that $\Phi(Q)[\beta, t] = p$, and $\Phi(Q)[\alpha, u] = r$, but this is the density of the event that $(\Phi(Q_\beta)[t], \Phi(Q_\alpha)[u]) = (p, r)$, which is equal to the integral of G_Φ over all pairs of temporal functions (P, R) such that $P[t] = p$ and $R[u] = r$. However, as we have already noted, $F_\Phi = G_\Phi$, implying that $f = g$. Apply Corollary 3.3 to complete the proof. ∎

The proof of II is similar.

The rest of Section 4 is devoted to showing how to construct two kinds of simple texture quilts. In Section 5, we apply these construction techniques in an experiment to investigate what sorts of textural characteristics are actually processed for motion information by the visual system.

4.4. *Binary Texture Quilts*

4.4.1. *A General Technique for Constructing Binary Texture Quilts.* The simplest sorts of texture quilts involve only two contrast values. As in Definition 4.1, let $T = \{t : t_0 \leq t < t_N\}$ be the temporal region of activity, with new timeblocks beginning at times $t_0, t_1, \ldots, t_N$. Let A be the spatial region of activity. Associate

with timeblocks $i = 0, 1, \ldots, N-1$ spatial functions f_i (called *timeblock pictures*), each of which is 0 everywhere outside A, and takes only the values 1 and -1 within A. In addition, associate with timeblocks 0 through $N-1$ a family

$$\phi_0, \phi_1, \ldots, \phi_{N-1} \tag{36}$$

of jointly independent random variables, each of which takes the value 1 or -1 with equal probability. Then, for $i = 0, 1, \ldots, N-1$, set

$$B_i[x, y, t] = \begin{cases} f_i[x, y] & \text{if } t \text{ is in timeblock } i, \\ 0 & \text{otherwise,} \end{cases} \tag{37}$$

and construct the random stimulus

$$B = \phi_0 B_0 + \phi_1 B_1 + \cdots + \phi_{N-1} B_{N-1}. \tag{38}$$

It is easy to see that B is a texture quilt. First, the functions B_i are defined to satisfy Definition 4.1(i) and (ii). The joint independence of the random variables ϕ_i ensures that B satisfies Definition 4.1(iii). To see that Definition 4.1(iv) is satisfied, note that for any $\alpha, \beta \in A$, either (i) $B_i[\alpha, t_i] = B_i[\beta, t_i]$ or (ii) $B_i[\alpha, t_i] = -B_i[\beta, t_i]$. In case (i),

$$B[\alpha, t_i] = \phi_i B_i[\alpha, t_i] = \phi_i B_i[\beta, t_i] = B[\beta, t_i], \tag{39}$$

implying that the pair $(B[\alpha, t_i], B[\beta, t_i])$ is distributed identically to the pair $(B[\beta, t_i], B[\alpha, t_i])$ (each pair with an equal probability of taking the value $(1, 1)$ or $(-1, -1)$). In case (ii)

$$B[\alpha, t_i] = -B[\beta, t_i], \tag{40}$$

and the pair $(B[\alpha, t_i], B[\beta, t_i])$ is distributed identically to the pair $(B[\beta, t_i], B[\alpha, t_i])$, each with an equal probability of assuming the value $(1, -1)$ or $(-1, 1)$. Thus Definition 4.1(iv) is satisfied along with 4.1(i), (ii), and (iii).

4.4.2. *Stimulus: The Sidestepping, Randomly Contrast-Reversing, Vertical Edge.* In Fig. 4b are displayed the 9 timeblock pictures comprising a particularly simple binary texture quilt. Note that the vertical dimension of Fig. 4b combines time and vertical space, precisely as a strip of movie film, scanned vertically, combines time and space. Timeblock pictures are separated by gray lines. Figure 4a shows the timeblock pictures f_0 through f_8 used in the construction. f_0 assigns the value -1 to all points (x, y) of the horizontal rectangle comprising the spatial region of activity, A. f_1 assigns 1 to the points in the leftmost eighth of A, and -1 to the points in the right seven-eighths. The timeblock pictures f_2 through f_8 continue to shift the vertical edge rightward through A until, in picture 8, A is uniformly 1. Multiplying each timeblock picture $i = 1, 2, \ldots, 9$ by its associated random variable ϕ_i yields, in this particular realization, the stimulus given in Fig. 4b.

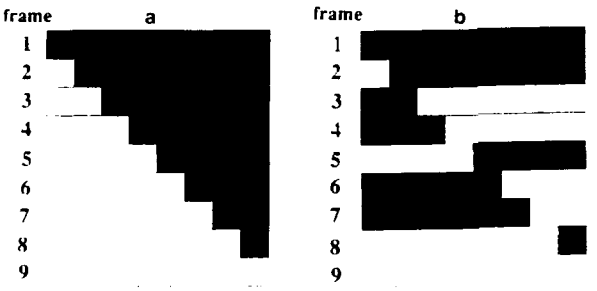


FIG. 4. Edge-driven motion from an ordinary edge and from a binary texture quilt. (a) A rightward moving light-dark edge visible to Fourier and non-Fourier motion systems. Nine entire frames are shown; each frame consists of an area of contrast +1 and area of contrast −1. (b) A realization of the *sidestepping, randomly contrast-reversing vertical edge*. This random stimulus is a texture quilt and hence microbalanced under all purely temporal transformations: that is, its rightward motion would be inaccessible to standard motion-analysis even if this analysis were preceded by an arbitrary, purely temporal transformation. Each frame of (b) was derived from the corresponding frame of (a) by multiplying the entire frame by a random variable that takes the value 1 or −1 with equal probability. The frame random variables are jointly independent. A straightforward way to extract the motion of this texture quilt is to (i) apply a linear filter sensitive to vertical edges, (ii) rectify the filtered output, and (iii) submit the result to standard motion analysis.

The construction of the sidestepping contrast-reversing edge (Fig. 4b) is symmetric to the construction of the traveling contrast-reversal of a random black-or-white vertical bar pattern (J in Fig. 2a). Transposing the x and t dimensions in Fig. 4b gives the xt-cross-section of a random stimulus J (e.g., Fig. 2a). This stimulus exhibits an unusual symmetry between space and time. Whereas the texture quilt of Fig. 4b is microbalanced under all purely temporal transformations, its transpose J (Fig. 2b) is microbalanced under all *purely spatial* transformations. Extracting motion from J requires *temporal* filtering followed by a nonlinearity. This process is essentially different from the process by which motion is extracted from texture quilts (e.g., Figs. 4b, 7a, 7b, and 7c) which requires a *spatial* nonlinearity.

4.4.3. *Stimulus: Oppositely Oriented Static Squarewaves Selected by a Drifting Grating.* Figure 5d shows the four timeblock pictures comprising another binary texture quilt constructed using technique 4.4.1. In Fig. 5a is shown a probabilistically defined sinewave grating, a stimulus whose motion is readily extracted by standard motion-analysis. In Figs. 5b1 and 5b2 are shown static vertical and horizontal squarewave gratings. The stimulus of Fig. 5c is obtained by using Fig. 5a to select between the vertical and horizontal gratings of Figs. 5b1 and 5b2. If the function of Fig. 5a is 1 at a certain point in space-time, the corresponding point in Fig. 5c is assigned the value of the corresponding point in Fig. 5b1; otherwise the point in Fig. 5c is assigned the value of the corresponding point in Fig. 5b2. Although Figs. 5c and 5d look similar, they differ in an important respect: the

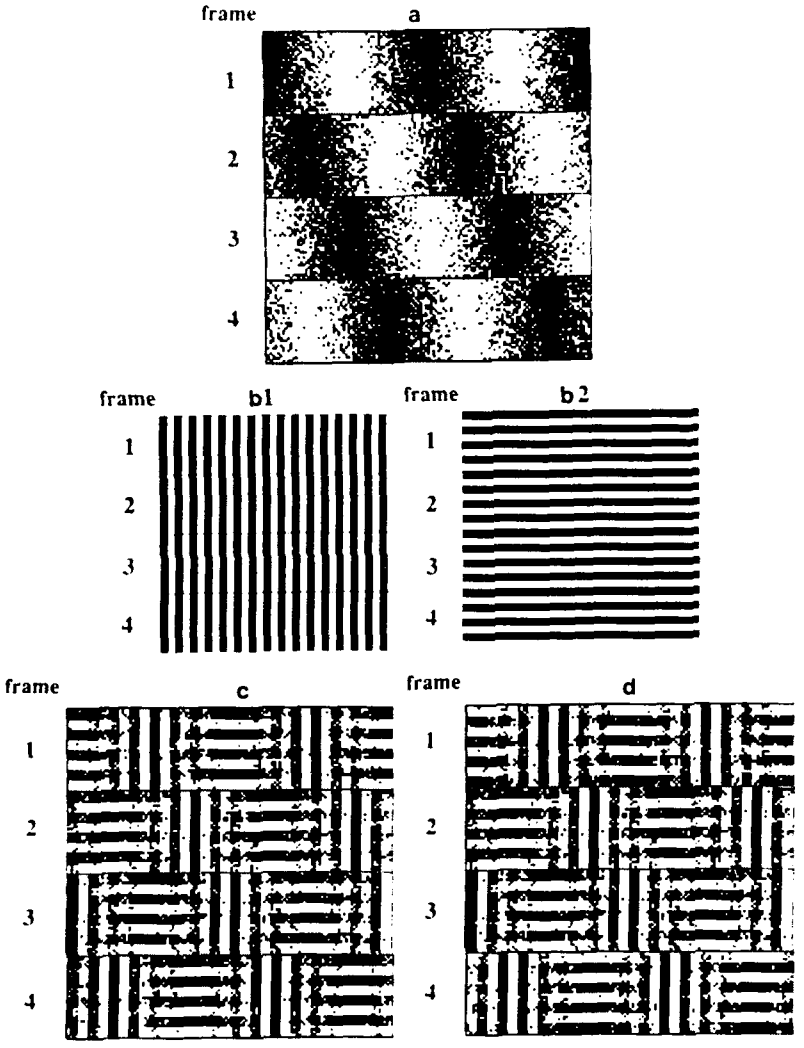


FIG. 5. Orientation-driven non-Fourier motion from a binary texture quilt. (a) A probabilistically defined sinewave grating that steps rightward 90 degrees between frames. The rightward motion in (a) is accessible to all motion detectors. (b1) Four frames of a static, vertical squarewave grating; (b2) Four frames of a static horizontal squarewave grating. (c) A rightward translating texture pattern. For every white point in (a), the corresponding value in (c) is chosen from the vertical squarewave grating in (b1); for every black point in (a), the corresponding value in (c) is chosen from the horizontal squarewave grating in (b2). (c) is not microbalanced; standard motion-analyzers can be designed to detect its motion. (d) A texture quilt. The frames of (d) are derived by multiplying the corresponding frames of (c) by jointly independent random variables, each of which takes the value 1 or −1 with equal probability. The texture quilt (d) is microbalanced under all purely temporal transformations, and therefore its rightward motion is unavailable to any mechanism that applies standard motion analysis to a purely temporal transformation of the visual signal.

stimulus of Fig. 5d is microbalanced under all purely temporal transformations, while that of Fig. 5c is not microbalanced. It is possible to design Fourier mechanisms to detect the motion of Fig. 5c, but not that of Fig. 5d. The critical difference is that the timeblock pictures of Fig. 5d are jointly independent, while those of Fig. 5c are not: Fig. 5d is obtained by randomly reversing the contrasts of the timeblock pictures of Fig. 5c.

4.5. *Sinusoidal Texture Quilts*

It is not difficult to elaborate technique 4.4.1 to a method for constructing quilts involving textures of arbitrarily many contrast values. We illustrate the principle in the construction of quilts comprised of patches of sinusoidal grating.

4.5.1. *A General Technique for Constructing Sinusoidal Texture Quilts.* As in Definition 4.1, let $T = \{t \mid t_0 \leqslant t < t_N\}$ be the temporal region of activity, with new timeblocks beginning at times $t_0, t_1, \ldots, t_{N-1}$. Let A be the spatial region of activity. Associate with timeblocks $i = 0, 1, \ldots, N-1$, spatial functions W_i, each of which is 0 everywhere outside A, and takes only the values 1 and -1 within A. The stimulus in each time block will be composed of two components characterized by spatial frequencies (ω_i, θ_i) and $(\tilde{\omega}_i, \tilde{\theta}_i)$, respectively, and independent phases $\rho_i, \tilde{\rho}_i$, respectively. Let

$$\omega_0, \theta_0, \tilde{\omega}_0, \tilde{\theta}_0, \omega_1, \theta_1, \tilde{\omega}_1, \tilde{\theta}_1, \ldots, \omega_{N-1}, \theta_{N-1}, \tilde{\omega}_{N-1}, \tilde{\theta}_{N-1} \tag{41}$$

be integers. Let P be an integer, and let

$$\rho_0, \tilde{\rho}_0, \rho_1, \tilde{\rho}_1, \ldots, \rho_{N-1}, \tilde{\rho}_{N-1} \tag{42}$$

be jointly independent random variables, each uniformly distributed on the set $\{0, 1, \ldots, P-1\}$. Then, define the stimulus S as the sum of N component stimuli S_i defined in each timeblock,

$$S = \sum_{i=0}^{N-1} S_i, \tag{43}$$

where, for $i = 0, 1, \ldots, N-1$, S_i is zero everywhere outside timeblock i; and for all t in timeblock i,

$$S[x, y, t] = I[x, y] = \begin{cases} \cos(2\pi(\omega_i x + \theta_i y - \rho_i)/P) & \text{if } W[x, y] = 1, \\ \cos(2\pi(\tilde{\omega}_i x + \tilde{\theta}_i y - \tilde{\rho}_i)/P) & \text{if } W[x, y] = -1, \\ 0 & \text{otherwise.} \end{cases} \tag{44}$$

It is easy to check that S satisfies Definition 4.1(i) and (ii). The joint independence of the random phase variables $\rho_i, \tilde{\rho}_i$, for $i = 0, 1, \ldots, N-1$ entails Definition 4.1(iii).

It remains to check that S satisfies Definition 4.1(iv). Consider points $\alpha, \beta \in A$. If $W_1[\alpha] \neq W_1[\beta]$, then, as is easily checked, $S[\alpha, t_1]$ and $S[\beta, t_1]$ are independent and identically distributed (each assuming a value from among $\{\cos(2\pi p/P) \quad p = 0, 1, \ldots, P-1\}$ with equal probability). On the other hand, if $W_1[\alpha] = W_1[\beta]$, then the pair $(S[\alpha, t_1], S[\beta, t_1])$ is distributed identically to the pair $(S[\beta, t_1], S[\alpha, t_1])$ as a consequence of the following

LEMMA. *Let $P \in \mathbf{Z}$, and let $\alpha = (\alpha_1, \alpha_2)$, $\beta = (\beta_1, \beta_2)$ and $\omega = (\omega_1, \omega_2)$ all be elements of $\mathbf{Z}^2$. Then for any integer $p \in \{0, 1, \ldots, P-1\}$, there exists an integer $q \in \{0, 1, \ldots, P-1\}$ such that (writing $\cdot$ for dot product)*

$$\cos(2\pi(\omega \cdot \alpha - p)/P) = \cos(2\pi(\omega \cdot \beta - q)/P) \tag{45}$$

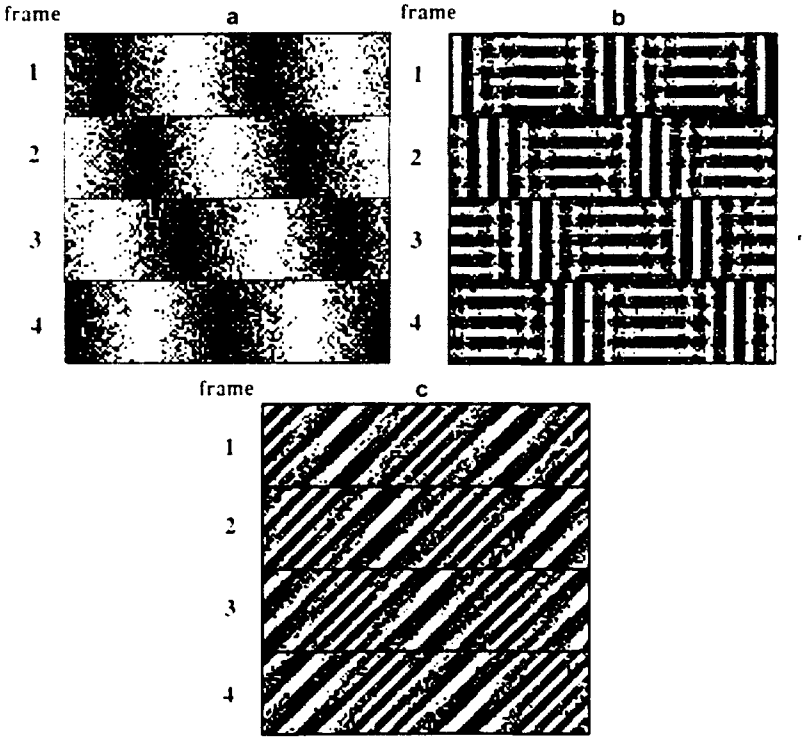


FIG. 6. Sinusoidal texture quilts: Motion driven by differences in *orientation* and in *spatial frequency*. (b) and (c) show realizations of random stimuli, each of which is microbalanced under all purely temporal transformations. Their rightward motion cannot be detected by any mechanism that applies standard motion analysis to a purely temporal transformation of the signal. In each case, the four frames in (a) select between two sinusoidal patterns. The phases of sinusoids are jointly independent across frames and across different-frequency sinusoidal components patched together in the same frame. The sinusoids mixed in (b) differ in orientation, whereas the sinusoids mixed in (c) have the same orientation, but differ in spatial frequency.

and

$$\cos(2\pi(\omega \cdot \beta - p)/P) = \cos(2\pi(\omega \cdot \alpha - q)/P). \tag{46}$$

Proof. As the reader may check, this is true for $q = (\omega \cdot \alpha + \omega \cdot \beta - p)$ modulo P. ∎

Thus, for α, β such that $W_i[\alpha] = W_i[\beta]$, we observe that for any outcome $\rho_i = p$, there exists an equally likely outcome $\rho_i = q$, such that

$$(\cos(2\pi(\omega_i \cdot \alpha - p)/P), \cos(2\pi(\omega_i \cdot \beta - p)/P)) = (\cos(2\pi(\omega_i \cdot \beta - q)/P), \cos(2\pi(\omega_i \cdot \alpha - q)/P)). \tag{47}$$

We infer that the pair $(S[\alpha, t_i], S[\beta, t_i])$ is distributed identically to the pair $(S[\beta, t_i], S[\alpha, t_i])$.

4.5.2. *Stimulus: Oppositely Oriented Static Sinusoids Selected by a Drifting Grating.* The sinusoidal analog to the binary texture quilt of Fig. 5d is shown in Fig. 6b. In Fig. 6a are shown the functions W_1, W_2, W_3, and W_4 used to select between horizontal and vertical gratings. For this quilt, $\tilde{\omega}_i = \tilde{\theta}_i = 0$, for $i = 1, 2, 3, 4$; and for some integer F (with F/P the number of cycles per pixel), $\omega_i = \tilde{\theta}_i = F$. The texture quilt of Fig. 6b modulates textural orientation across space and time. Alternatively, we can just as easily keep orientation constant and vary spatial frequency.

4.5.3. *Stimulus: Static Sinusoids of Different Spatial Frequencies, Selected by a Drifting Grating.* Figure 6c shows a texture quilt using the sampling functions of Fig. 6a, but setting $\omega_i = \theta_i = 2\tilde{\omega}_i = 2\tilde{\theta}_i$ for $i = 1, 2, \ldots, 4$.

5. What Aspects of Texture Does the Visual System Process for Motion?

In this section, we describe a psychophysical experiment investigating the question of what characteristics of spatial texture are analyzed for motion information by the visual system. Three texture quilts are compared across four different viewing conditions. These conditions comprise a sequence of similar but increasingly challenging motion discrimination tasks.

5.1. *Procedure*

Every texture quilt used in this experiment is comprised of a sequence of jointly independent timeblocks, each lasting 1/30 s. (Each timeblock consists of two identical refreshes at 1/60 s.) Each texture quilt is stochastically periodic with a period of 8 timeblocks; that is, for any integer i, the ith timeblock is identically distributed to the $i + 8$th timeblock. Accordingly, we refer to 8 timeblocks of the texture quilt as one *cycle*. The motion elicited by each quilt is carried by a squarewave that selects between two textures, and steps 1/4 cycle on every odd timeblock. The squarewave thus completes one of its four-step cycles in each 8 timeblock cycle of the quilt.

On each trial, a texture quilt moving randomly left or right is presented, and the subject is required to signal (with a button-press) which way the quilt appeared to move. The subject is asked to maintain fixation on a small spot present in the middle of the stimulus throughout the display, and receives feedback after each trial. For each quilt under each viewing condition, the subject performs 100 practice trials followed directly by 100 actual trials. Quilt realizations are jointly independent across trials. The starting phase of the quilt is chosen randomly on each trial.

The Four Viewing Conditions. For a given quilt, the four viewing conditions differ with respect to the number of quilt cycles displayed. In Condition 1, the easiest condition, the subject sees two quilt cycles (each cycle comprised of eight stimulus timeblocks), with each timeblock displayed for 1.30 s. In Conditions 2, 3, and 4, the subject sees 1.5, 1, and 0.5 quilt cycles, respectively.

5.1.1. *Three Quilt Stimuli.* The first quilt (the **F**-quilt) modulates textural spatial frequency as a function of space and time, while keeping orientation constant. The 8 timeblocks comprising one full cycle of the **F**-quilt are shown in Fig. 7a. A second quilt (the **O**-quilt, Fig. 7b) modulates textural orientation as a function of space and time, while keeping spatial frequency constant. A third quilt (the **E**-quilt, Fig. 7c) spatiotemporally modulates texture between jointly independent binary noise and the so-called "even" texture (Julesz, Gilbert, & Victor, 1978).

All stimuli were viewed from 1 m against a mean luminant background. At this distance, each quilt spanned 6.8 horizontal and 3.2 vertical degrees, and the modulating squarewave moved at an average velocity of 12.75 deg s.

5.1.2. *Why These Three Quilts.* In each of the three quilts, a squarewave with vertical bars is used to modulate between two textures as a function of space and time. The squarewave has a spatial frequency of 0.3 c deg, and steps 1/4 cycle rightward on every odd timeblock (temporal frequency 3.75 Hz, velocity 12.75 deg s). We use a 1/4 cycle stepping squarewave to modulate between the two textures comprising each quilt in order to rule out the possibility that the motion elicited by the quilt is being carried by the border between textural regions. That is, the 1/4 cycle stepping squarewave has the advantage that the signal derived from the borders between texture regions is ambiguous in motion content. Given the requirement of 1/4 cycle steps, we changed the particular instantiation of the quilt on even timeblocks (i.e., within steps of the squarewave) in order to spread textural energy broadly in temporal frequency without altering the spatial frequency content of the texture.

It has been previously observed (Green, 1986; Ramachandran, Ginsburg, & Anstis, 1983; Watson & Ahumada, 1983a) that motion is carried more effectively by spatiotemporal variation of textural spatial frequency than by variation of textural orientation. The **F**-quilt and **O**-quilt were chosen to further investigate this claim. The **E**-quilt is of interest because the two textures of which it is composed (jointly independent binary noise and the even texture) have identical second order

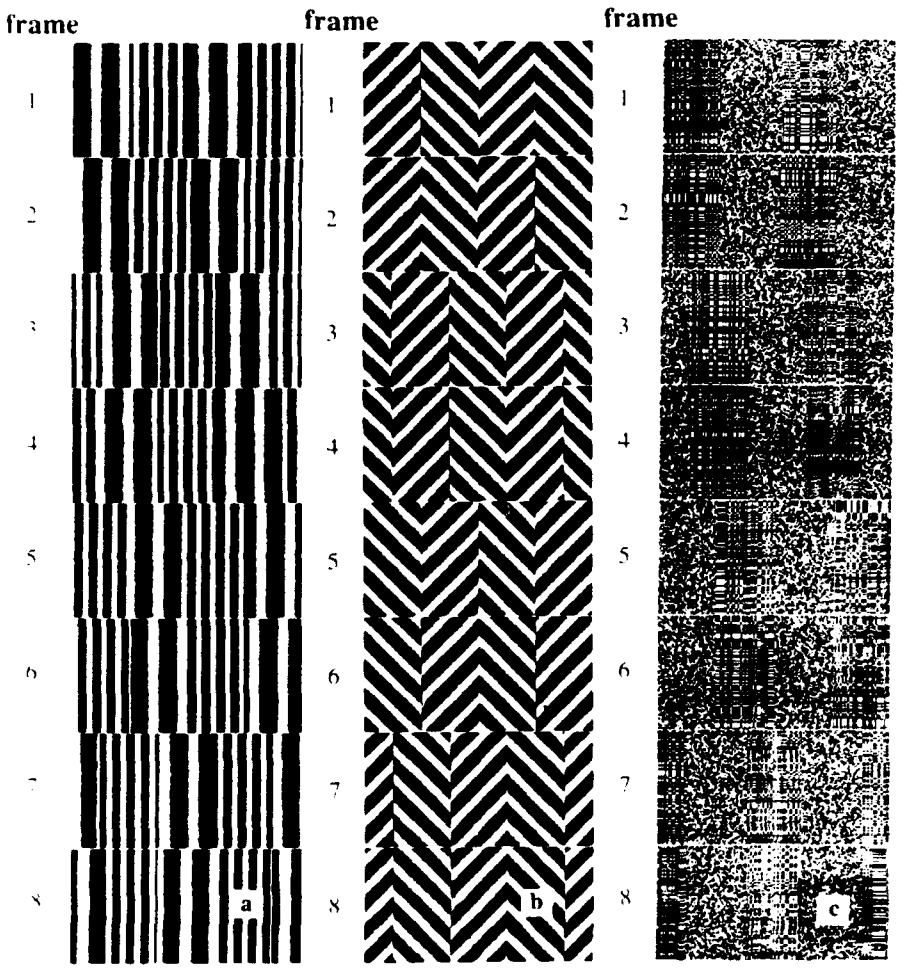


FIG. 7. Three quilts used to study motion carried by modulation of texture spatial frequency, by texture orientation, and by higher order textural characteristics. (a) Eight frames that comprise one cycle of the **F**-quilt. Motion is generated by a squarewave modulation of textural spatial frequency. The squarewave grating selects between vertical sinusoidal gratings of spatial frequency 1.2 and 2.4 c deg. The texture-modulating squarewave is 0.3 c deg, and steps 1.4 cycle rightward on every odd frame. Every even frame is independent of and distributed identically to the preceding frame. Presentation proceeds at the rate of 30 frames s. This gives the texture-modulating squarewave a temporal frequency of 3.75 Hz and a mean velocity of 2.5 deg s. (b) Eight frames that comprise one cycle of the **O**-quilt. In the **O**-quilt, textural orientation is modulated by the same squarewave used to modulate spatial frequency in the **F**-quilt. The **O**-quilt squarewave selects between oppositely oriented sinusoidal gratings that have a spatial frequency of 2.8 c deg. (c) Eight frames that comprise one cycle of the **E**-quilt. In the **E**-quilt, the texture-modulating squarewave selects between jointly independent binary noise and an even texture (Julesz, Gilbert, & Victor, 1978). Despite the evident difference between these two textures, every time-independent linear filter has the same expected power for both textures. Thus, if motion-from-texture resulted from applying a simple squaring transformation to the output of a spatial linear filter and submitting the result to standard motion analysis, the motion of the **E**-quilt would be invisible.

statistics. That is, the joint distribution of any given pair of points in space is the same under both the component textures of the E-quilt. This means that, despite the obvious difference in appearance between the component textures, the expected energy in the response of any given spatial linear filter is the same for both component textures. If the pointwise nonlinearity applied to the output of the spatial linear filter prior to motion analysis were simple squaring, it would be impossible to detect the motion of the E-quilt.

Victor and Conte (1990) studied apparent motion elicited by E-quilts, and noted that it is much weaker than motion elicited by comparable stimuli (also texture quilts) that modulate between textures differing in spatial frequency. Our experiment confirms this finding.

5.2. *Results*

Two subjects participated in the study, CC (the experimenter) and GA (naive). The results for CC are shown in Fig. 8 bottom, and those for GA are shown in

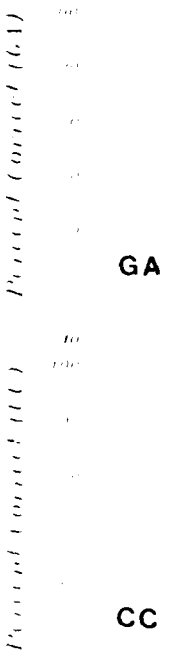


FIG. 8. The percent of correct direction-of-motion judgments to the E-quilt, the O-quilt, and the E-quilt as a function of stimulus duration. The panels show data for subjects CC and GA, respectively. Each data point is the mean of 100 judgments. (Squares) E-quilt, (triangles) O-quilt, (circles) E-quilt. The stimulus durations of 133, 266, 400, and 533 ms, correspond to stimulus presentations of 0.5, 1, 1.5, and 2 quilt cycles.

Fig. 8 top. Note first that both subjects were able to reliably discriminate left right motion in all three stimuli although subject GA failed with the **E**-quilt at the briefest exposure. The two subjects performed comparably well at motion direction discrimination of the **O**-quilt, but CC was much better than GA at detecting the motion of both the **F**-quilt and the **E**-quilt. Subject CC was better at detecting the motion of the **F**-quilt than the **O**-quilt; the reverse was true of subject GA.

It is possible that these performance differences reflect a genuine differences in the perceptual apparatus of the two subjects. However, we cannot rule out the possibility that the better performance of subject CC is due merely to his vastly greater experience with motion perception tasks of this sort.

5.3. *Discussion*

Many of the models proposed to explain rapid, preattentive segregation of spatial textures (Beck, Sutter, & Ivry, 1987; Bergen & Adelson, 1988; Caelli, 1985; Malik & Perona, 1989; Sutter, Beck, & Graham, 1989) can easily be adapted to deal with the motion displayed by texture quilts. The texture segregation models in this class typically subject the visual input function to a linear transformation (a "texture grabber") followed by a pointwise nonlinearity (such as a rectifier or thresholder) to indicate the presence or absence of the texture. Such models propose that two contiguous textural regions would generate a perceptual boundary if the visual system were equipped with a linear filter that is differentially tuned to one of the textures.

An analogous mechanism to detect the motion of texture quilts, suggested by the current experiment and the work of Victor and Conte (1990), (i) convolves the input stimulus with a spatial texture-grabbing filter tuned to the moving texture, then (ii) squares the output of the filter, to transform regions of high energy filter output into regions of high average value, and (iii) subjects the rectified output to standard motion analysis. However, the transformation applied in steps (i) and (ii) does not distinguish between the two textures comprising the **E**-quilt, and therefore fails to account for the good performance with the **E**-quilt. A simple modification to deal with texture segregation and motion perception of the **E**-quilt is to assume some other post-filter rectification operation than the squaring operation. It is quite easy to choose a linear filter in combination with a post-filter rectifier (other than the squaring operation) that will segregate the random and even textures (e.g., Julesz & Bergen, 1983). The current experiment does not specifically indicate the kind of rectification that might be involved.

What sorts of filters are available to the visual system to compute motion from texture? For example, Daugman (1985) points out that (i) Gabor filters provide an optimal tradeoff between resolution in the space and spatial frequency domains, and (ii) many investigators note that simple cells in cat striate cortex are well modeled by oriented Gabor filters (e.g., Andrews & Pollen, 1979; DeValois, DeValois, & Yund, 1979; Wilson & Sherman, 1976). Are the linear filters that serve motion-from-texture computations Gabor-like cortical simple cells? The theory

reported here provides a tool, and the demonstration experiments illustrate how it might be used to answer such questions.

6. Summary

The main contributions of this paper are to (i) introduce the notion of a random stimulus *microbalanced under all pointwise transformations*, (ii) provide necessary and sufficient conditions for a random stimulus to be of this sort, (iii) use this result to construct apparent motion stimuli called *texture quilts* that are microbalanced under all purely temporal transformations, and (iv) show that subjects can reliably discriminate the motion direction of three kinds of texture quilts.

Texture quilts provide a flexible array of tools for studying motion perception that is truly mediated by spatiotemporal modulation of spatial texture without contamination by mechanisms responsive to the motion extracted directly by standard analysis or motion extracted by standard analysis of any purely temporal transformation of the stimulus.

Acknowledgments

The research reported here was supported by USAF Life Science Directorate, Visual Information Processing Program Grants 85-0364 and 88-0140.

References

Adelson, E. H., & Bergen, J. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A*, **2**(2), 284–299.

Adelson, E. H., & Bergen, J. (1986). The extraction of spatio-temporal energy in human and machine vision. *Proceedings of the IEEE Workshop on Motion: Representation and Analysis*, 151–155.

Andrews, B. W., & Pollen, D. A. (1979). Relationship between spatial frequency selectivity and receptive field profile of simple cells. *Journal of Physiology (London)*, **287**, 163–176.

Anstis, S. M. (1970). Phi movement as a subtraction process. *Vision Research*, **10**, 1411–1430.

Baker, C. L., & Braddick, O. (1982a). Does segregation of differently moving areas depend on relative or absolute displacement. *Vision Research*, **22**, 851–856.

Baker, C. L., & Braddick, O. (1982b). The basis of area and dot number effects in random dot motion perception. *Vision Research*, **22**, 1253–1260.

Beck, J., Sutter, A., & Ivry, R. (1987). Spatial frequency channels and perceptual grouping in texture segregation. *Computer Vision, Graphics, and Image Processing*, **37**, 299–325.

Bell, H. H., & Lappin, J. S. (1979). The detection of rotation in random dot patterns. *Perception and Psychophysics*, **26**, 415–417.

Bergen, J. R., & Adelson, E. H. (1988). Early vision and texture perception. *Nature*, **333**(6171), 363–364.

Bowne, S. F., McKee, S. P., & Glaser, D. A. (1989). Motion interference in speed discrimination. *Journal of the Optical Society of America A*, **6**(7), 1112–1121.

Braddick, O. (1973). The masking of apparent motion in random-dot patterns. *Vision Research*, **13**, 355–359.

BRADDICK, O. (1974). A short-range process in apparent motion. *Vision Research*, **14**, 519–527.

CAELLI, T. (1985). Three processing characteristics of visual texture segmentation. *Spatial Vision*, **1**(1), 19–30.

CAVANAGH, P. (1988). Motion: The long and the short of it. Presented at *Conference on Visual Form and Motion Perception: Psychophysics, Computation, and Neural Networks* (Meeting dedicated to the memory of the late Kvetoslav Prazdny). Boston University, MA, March 5, 1988.

CAVANAGH, P., ARGUIN, M., & VON GRUNAU, M. (1989). Interattribute apparent motion. *Vision Research*, **29**(9), 1197–1204.

CHANG, J. J., & JULESZ, B. (1983a). Displacement limits, directional anisotropy and direction versus form discrimination in random dot cinematograms. *Vision Research*, **23**, 639–646.

CHANG, J. J., & JULESZ, B. (1983b). Displacement limits for spatial frequency random-dot cinematograms in apparent motion. *Vision Research*, **23**, 1379–1386.

CHANG, J. J., & JULESZ, B. (1985). Cooperative and non-cooperative processes of apparent movement of random dot cinematograms. *Spatial Vision*, **1**(1), 39–45.

CHUBB, C., & SPERLING, G. (1987). Drift-balanced random stimuli: A general basis for studying non-Fourier motion perception. *Investigative Ophthalmology and Visual Science*, **28**, 233.

CHUBB, C., & SPERLING, G. (1988). Drift-balanced random stimuli: A general basis for studying non-Fourier motion perception. *Journal of the Optical Society of America A*, **5**(11), 1986–2007.

DAUGMAN, J. G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America A*, **2**(7), 1160–1169.

DERRINGTON, A. M., & BADCOCK, D. R. (1985). Separate detectors for simple and complex grating patterns? *Vision Research*, **25**, 1869–1878.

DERRINGTON, A. M., & HENNING, G. B. (1987). Errors in direction-of-motion discrimination with complex stimuli. *Vision Research*, **27**, 61–75.

DEVALOIS, K. K., DEVALOIS, R. L., & YUND, E. W. (1979). Responses of striate cortical cells to grating and checkerboard patterns. *Journal of Physiology (London)*, **291**, 483–505.

VAN DOORN, A. J., & KOENDERINK, J. J. (1984). Spatiotemporal integration in the detection of coherent motion. *Vision Research*, **24**, 47–54.

DOSHER, BARBARA A., LANDY, M. S., & SPERLING, G. (1989). Ratings of kinetic depth in multi-dot displays. *Journal of Experimental Psychology: Human Perception and Performance*, **15**, 116–425.

GREEN, M. (1986). What determines correspondence strength in apparent motion. *Vision Research*, **26**, 599–607.

JULESZ, B. (1971). *Foundations of cyclopean perception*. Chicago: Univ. of Chicago Press.

JULESZ, B., & BERGEN, J. R. (1983). Textons, the fundamental elements in preattentive vision and perception of textures. *Bell Systems Technical Journal*, **62**(6), 1619–1645.

JULESZ, B., GILBERT, E., & VICTOR, J. D. (1978). Visual discrimination of textures with identical third-order statistics. *Biological Cybernetics*, **31**, 137–140.

LAPPIN, J. S. (1989). Personal communication, June 20.

LAPPIN, J. S., & BELL, H. H. (1972). Perceptual differentiation of sequential visual patterns. *Perception and Psychophysics*, **12**, 129–134.

LEEKINS, A. M. M., & KOENDERINK, J. J. (1984). Illusory motion in visual displays. *Vision Research*, **24**, 1083–1090.

MALIK, J. & PERONA, P. (1989). *A computational model of texture perception* (Computer Science Division (EECS) Report No. UCB CSD 89-491). Berkeley: University of California.

MARR, D. & ULLMAN, S. (1981). Direction selectivity and its use in early visual processing. *Proc. R. Soc. London, Ser. B*, **211**, 151–180.

NAKAYAMA, K., & SILVERMAN, G. (1984). Temporal and spatial characteristics of the upper displacement limit for motion in random dots. *Vision Research*, **24**, 293–300.

PANTLE, A. & TURANO, K. (1986). Direct comparisons of apparent motions produced with luminance, contrast-modulated (CM), and texture gratings. *Investigative Ophthalmology and Visual Science*, **27**(3), 141.

PETERSIK, J. T., HICKS, K. L., & PANTLE, A. J. (1978). Apparent movement of successively generated subjective figures. *Perception*, **7**, 371–383.

RAMACHANDRAN, V. S., & ANSTIS, S. M. (1983). Displacement thresholds for coherent apparent motion in random dot-patterns. *Vision Research*, **23**, 1719–1724.

RAMACHANDRAN, V. S., GINSBURG, A., & ANSTIS, S. M. (1983). Low spatial frequencies dominate apparent motion. *Perception*, **12**, 457–461.

RAMACHANDRAN, V. S., RAO, V. M., & VIDYASAGAR, T. R. (1973). Apparent movement with subjective contours. *Vision Research*, **13**, 1399–1401.

VAN SANTEN, J. P. H., & SPERLING, G. (1984). A temporal covariance model of motion perception. *Journal of the Optical Society of America A*, **1**, 451–473.

VAN SANTEN, J. P. H., & SPERLING, G. (1985). Elaborated Reichardt detectors. *Journal of the Optical Society of America A*, **2**(2), 300–321.

SNAPLEY, R., & ENROTH-CUGELL, C. (1984). Visual adaptation and retinal gain controls. *Progress in Retinal Research*, **3**, 263–346, 1984.

SPERLING, G. (1976). Movement perception in computer-driven visual displays. *Behavior Research Methods and Instrumentation*, **8**, 144–151.

SUTTER, A., BECK, J., & GRAHAM, N. (1989). Contrast and spatial variables in texture segregation: Testing a simple spatial-frequency channels model. *Perception and Psychophysics*, **46**(4), 312–332.

TURANO, K., & PANTLE, A. (1989). On the mechanism that encodes the movement of contrast variations: Velocity discrimination. *Vision Research*, **29**(2), 207–221.

ULLMAN, S. (1979). *The interpretation of Visual Motion*. Cambridge, MA: MIT Press.

VICTOR, J. D., & CONTE, M. M. (1990). Motion mechanisms have only limited access to form information. *Vision Research*, **30**(2), 289–301.

WATSON, A. B., & AHUMADA, A. J., JR. (1983a). A linear motion sensor. *Perception*, **12**, A17.

WATSON, A. B., & AHUMADA, A. J., JR. (1983b). *A look at motion in the frequency domain*. NASA Technical Memorandum 84352.

WATSON, A. B., & AHUMADA, A. J., JR. (1985). A model of human visual-motion sensing. *Journal of the Optical Society of America A*, **2**(2), 322–342.

WATSON, A. B., AHUMADA, A. J., & FARRELL, J. E. (1986). The window of visibility: A psychophysical theory of fidelity in time-sampled motion displays. *Journal of the Optical Society of America A*, **3**(3), 300–307.

WILSON, J., & SHERMAN, S. (1976). Receptive field characteristics of neurones in cat striate cortex: Changes with visual field eccentricity. *Journal of Neurophysiology*, **39**, 512–533.

RECEIVED: February 8, 1989

Printed by Catherine Press, Ltd., Tempelhof 41, B-8000 Brugge, Belgium

Vision Res. Vol. 31, No. 5, pp. 859–876, 1991
Printed in Great Britain

0042-6989/91 $3.00 + 0.00

THE KINETIC DEPTH EFFECT AND OPTIC FLOW—II. FIRST- AND SECOND-ORDER MOTION

MICHAEL S. LANDY,[1] BARBARA A. DOSHER,[2] GEORGE SPERLING[1] and MARK E. PERKINS[1]

[1]Psychology Department, New York University, NY 10003 and [2]Psychology Department, Columbia University, NY 10027, U.S.A.

(*Received* 24 *August* 1989; *in revised form* 1 *May* 1990)

Abstract—We use a difficult shape identification task to analyze how humans extract 3D surface structure from dynamic 2D stimuli—the kinetic depth effect (KDE). Stimuli composed of luminous tokens moving on a less luminous background yield accurate 3D shape identification regardless of the particular token used (either dots, lines, or disks). These displays stimulate both the 1st-order (Fourier-energy) motion detectors and 2nd-order (nonFourier) motion detectors. To determine which system supports KDE, we employ stimulus manipulations that weaken or distort 1st-order motion energy (e.g. frame-to-frame alternation of the contrast polarity of tokens) and manipulations that create *microbalanced* stimuli which have no useful 1st-order motion energy. All manipulations that impair 1st-order motion energy correspondingly impair 3D shape identification. In certain cases, 2nd-order motion could support limited KDE, but it was not robust and was of low spatial resolution. We conclude that 1st-order motion detectors are the primary input to the kinetic depth system. To determine minimal conditions for KDE, we use a two frame display. Under optimal conditions, KDE supports shape identification performance at 63–94% of full-rotation displays (where baseline is 5%). Increasing the amount of 3D rotation portrayed or introducing a blank inter-stimulus interval impairs performance. Together, our results confirm that the human KDE computation of surface shape uses a global optic flow computed primarily by 1st-order motion detectors with minor 2nd-order inputs. Accurate 3D shape identification requires only two views and therefore does not require knowledge of acceleration.

KDE Kinetic depth effect Structure from motion Shape Optic flow

INTRODUCTION

When a collection of randomly positioned dots moves on a CRT screen with motion paths that are projections of rigid 3D motion, a human viewer perceives a striking impression of three-dimensionality and depth. This phenomenon of depth computed from relative motion cues is known as the kinetic depth effect (KDE: Wallach & O'Connell, 1953).

What are the important cues that lead to a 3D percept from such a display? Is it motion, or are there other important cues? If it is motion, then what kind of motion detection system(s) are used to support the structure-from-motion computation? Is a computation of velocity sufficient, or are more elaborate measurements necessary, such as of acceleration? These are the questions that we address in this paper.

In a series of recent papers (Dosher, Landy & Sperling, 1989a, b; Sperling, Landy, Dosher & Perkins, 1989; Sperling, Dosher & Landy, 1990), we examined the cues necessary for subjects to perceive an accurate representation of a 3D surface portrayed using random dot displays. In each trial of a new shape identification task we devised, subjects view a random dot representation of one of a set of 53 3D shapes and identify the shape and rotation direction. Shape identity feedback optimizes the subject's ability to compute shape from each type of motion stimulus. For accurate performance, the task requires either a 3D percept or a subject strategy that uses 2D velocity information in a manner that is computationally equivalent to that required to solve for 3D shape (Sperling et al., 1989, 1990; see the discussion of expt 2, below).

We have shown that the only cue used for the perception of three-dimensionality in these displays is motion (Sperling et al., 1989, 1990). Further experiments determined that global optic flow is used rather than the position information for individual dots, since accuracy remains high when dot lifetimes are reduced to as little as two frames (Dosher et al., 1989b). In that paper, we concluded that the input to the KDE computation is an optic flow generated by a 1st-order motion detection mechanism, such

as the Reichardt detector (Reichardt, 1957). Two manipulations that perturb 1st-order motion energy mechanisms—flicker and polarity alternation—also interfered with KDE (Dosher et al., 1989b). In polarity alternation, dots change over time from black to white to black on a gray background. When compared to dots that remain white, polarity alternation was equally or slightly more detectable in a detection task, was poorer but still well above chance in a discrimination of direction of motion task (computed, presumably, using tracking of the dots or using more elaborate, 2nd-order motion detection mechanisms) but was useless for tasks requiring KDE or motion segregation. These latter two tasks require the evaluation of velocity in a number of locations simultaneously (Sperling et al., 1989). Shape identification performance in a range of conditions was shown to be monotonic with a computed index of 1st-order net directional power in the stimuli (Dosher et al., 1989b). Hence, for sparse dot stimuli, KDE depends upon a simple spatio-temporal (1st-order) Fourier analysis of multiple local areas of the stimulus.

In this paper, we further examine and generalize the contributions of several types of motion detectors to the optic flow computations used by the structure-from-motion mechanism.

MOTION ANALYSIS MODELS AND THE KDE

1st-order motion analysis

To motivate the stimulus conditions studied here, we begin by summarizing models of early motion detection and analysis. Several recent motion detection models (van Santen & Sperling, 1984, 1985; Adelson & Bergen, 1985; Watson & Ahumada, 1985) share as a common antecedent the model proposed by Reichardt (1957). We refer to this class of models as 1st-order motion detectors. Below, 2nd-order mechanisms involving additional processing stages will be discussed. In the Reichardt detector, luminance is measured at two spatial locations A and B. The measurement at position A is delayed in time, and then cross-correlated over time with the measurement at position B, resulting in a "half-detector" sensitive to motion from position A to B. A second such "half-detector" sensitive to motion from B to A is set in opponency with the first, resulting in the full motion detector. van Santen and Sperling (1984, 1985) have investigated this model along with extensions involving voting rules for combining outputs of many detectors to enable predictions of psychophysical experiments, resulting in their Elaborated Reichardt Detector (ERD).

An alternative way of characterizing motion detection is in the frequency domain. A motion detector can be built of several linear spatio-temporal filters. Each filter is sensitive only to energy in two of the four quadrants in spatio-temporal Fourier space (ω_x, ω_t). In other words, the filters are not *separable*. Their receptive fields are oriented in space-time, and thus they are sensitive to motion in a particular direction and at a particular scale (Adelson & Bergen, 1985; Burr, Ross & Morrone, 1986; Watson & Ahumada, 1985). The Fourier "energy" (the squared output of a quadrature pair of filters) in each of two opposing motion directions is computed, and put in opponency. This "motion energy detector", proposed by Adelson and Bergen (1985), and the ERD differ in their construction and in the signals available at the subunit level, but are indistinguishable at their outputs (Adelson & Bergen, 1985; van Santen & Sperling, 1985).

The structure-from-motion computation relies upon the measurement of image velocities at several image locations. The KDE shape identification task that we use here can be solved by categorizing velocity at six spatial locations into three categories: leftward, approximately zero, and rightward (Sperling et al., 1989). Thus, in order to discriminate the 53 test shapes by KDE, motion detection must be followed by at least some rudimentary local velocity calculation.

In order to signal velocity, the outputs of more than one such 1st-order motion detector must be pooled. Speed may be computed by pooling only two detectors (a motion and a "static" detector, Adelson & Bergen, 1985). To signal motion direction, signals must be pooled across a variety of orientations (Watson & Ahumada, 1985). Finally, in order to solve the "aperture problem" for more complex stimuli (Burt & Sperling, 1981; Marr & Ullman, 1981), signals may be pooled over a variety of directions and perhaps scales (Heeger, 1987).

In the previous paper (Dosher et al., 1989b), shape identification performance was shown to relate directly to the quality of the signal available from 1st-order motion detection mechanisms. Each stimulus consisted of a large number of dots on a gray background representing a 2D projection of dots on the surface of a smooth 3D

shape under rotary oscillation. In one condition (contrast polarity alternation), the dots were first brighter than the background ("white-on-gray"), then darker than the background ("black-on-gray"), then bright again, in successive frames. For a dense random dot field (50% black/50% white) under simple planar motion, polarity alternation causes a percept of motion opposite to the true direction of motion (the "reverse-phi phenomenon", Anstis & Rogers, 1975); reverse-phi is thought to reflect a spatio-temporal Fourier analysis of the stimulus, since contrast reversal reverses the direction of motion of the lowest-frequency Fourier components (van Santen & Sperling, 1984). With contrast reversal, the outputs of 1st-order motion detection mechanisms no longer simply signal the intended direction and velocity of motion. Contrast reversal stimuli do not yield a depth-from-motion percept (Dosher et al., 1989b). We take this as evidence that the KDE relies upon input from a 1st-order motion analysis.

2nd-order motion analysis

For the sparse random dot stimuli (Dosher et al., 1989b), contrast polarity alternation eliminated the perception of structure from motion. Nonetheless, subjects could judge the direction of patches of contrast polarity alternating dots undergoing simple translation. What kind of a motion detector might be used to correctly judge the motion of a translating, polarity-alternating dot? One simple possibility would be to first apply a luminance nonlinearity to the input stimulus. For example, if the input stimulus were full-wave rectified about the mean luminance, the polarity-alternating stimulus would be converted to the equivalent of rigid motion of a white dot on a gray background. Thus, a full-wave rectifier of contrast followed by a 1st-order analyzer (such as those discussed above) would be capable of analyzing such a motion stimulus correctly (Chubb & Sperling, 1988b, 1989a, b).

A motion detection system consisting of a contrast nonlinearity followed by a 1st-order detector is one example of a wide class of "2nd-order detection mechanisms", each of which consists of a linear filtering of the input (spatial and or temporal), followed by a contrast nonlinearity, followed by a standard 1st-order motion detection mechanism. A number of results demonstrate the existence of both 1st- and 2nd-order motion mechanisms and show the contribution of both to the perception of planar motion (Anstis & Rogers, 1975; Chubb & Sperling, 1988b, 1989a, b; Lelkens & Koenderink, 1984; Ramachandran, Rao & Vidyasagar, 1973; Sperling, 1976).

Can both 1st- and 2nd-order motion mechanisms be used by the KDE system? The polarity-alternating dots did not yield an effective KDE percept of our 3D shapes. If one accepts the existence of both 1st- and 2nd-order motion mechanisms, why didn't the 2nd-order system support KDE? The KDE stimuli were relatively small (3.7 × 4.2 deg) and viewed foveally (eye movements were permitted throughout the 2 sec stimulus duration). Evidence from studies of planar motion suggests that both systems were available under these conditions (Chubb & Sperling, 1988b). For polarity alternation stimuli, the most salient low frequency components from the 1st-order system were in the wrong direction. We assume that the 2nd-order system yields a correct (if attenuated) analysis. Bad shape identification performance may have resulted either from the perturbed 1st-order analysis or because of competition between the 1st- and 2nd-order systems (which signaled opposite directions of motion in some frequency bands). Our evidence (Dosher et al., 1989b) demonstrated that 1st-order system input is the predominant input to KDE, but it did not exclude the possibility of input from 2nd-order motion detection mechanisms. To approach that question, we consider a KDE stimulus that produces a simple 2nd-order motion analysis, but to which the 1st-order motion system is, statistically, blind.

Microbalanced motion stimuli

Chubb and Sperling (1988b) defined a class of stimuli, called *microbalanced*, among which are stimuli with the properties that we desire. In expt 1 we concentrate on two examples of microbalanced motion stimuli. These stimuli are random in the sense that any given stimulus is a realization of a random process. As proven by Chubb and Sperling (1988b), if a stimulus is microbalanced then the expected output of every 1st-order detector (ERD or motion energy detector) will be zero. Thus, Chubb and Sperling defined a class of stimuli for which a consistent motion signal requires a 2nd-order motion analysis, and showed that the 2nd-order analysis predicted observers' percepts for several examples of the class

The polarity alternation stimulus is not microbal. .ced; any given frequency band does show consistent motion, with the lowest spatial frequencies signalling motion in the wrong direction. This stimulus can be transformed into a microbalanced one as follows: for each dot, choose the contrast polarity randomly and independently for every frame. Any given 1st-order detector will be just as likely to signal rightward motion as it is to signal leftward motion since it will either see the same contrast polarity across any successive pair of frames or it will see contrast polarity alternate, with equal probability. One question we examine in this paper is whether the motion signal available from 2nd-order mechanisms can be used to compute 3D structure.

We present two experiments. In the first, we examine performance on a shape identification task for a variety of KDE stimuli. Several types of stimuli provide good 1st-order motion. Others are microbalanced and hence can only be analyzed by 2nd-order mechanisms. Still others offer good 1st-order motion, but involve camouflage similar to that available in some of the microbalanced conditions. We find that 1st-order motion is used, and that input from 2nd-order mechanisms may also be used but is not as robust. In a second experiment, we examine the residual shape percept from two-frame KDE stimuli in order to determine whether a single velocity field is a sufficient cue for shape identification or whether acceleration also is needed.

EXPERIMENT 1. POLARITY ALTERNATION, MICROBALANCE, AND CAMOUFLAGE

In the first experiment, a shape discrimination task is used with a variety of displays. First, in order to sensibly compare results to our previous work (Sperling et al., 1989; Dosher et al., 1989b), there are control conditions that are identical to those of our previous experiments (the "Motion without density cue, standard speed, standard intensity" and "Motion with polarity alternation, standard speed, standard intensity" conditions of the preceding paper). In addition to dots, randomly positioned disks and lines are also used here in order to examine the effects of the foreground token used to carry the motion. The disk and line tokens are larger than the single pixel dots, and hence have more contrast energy. They enable us to test whether our previous failure to find KDE with polarity alternation resulted from the low contrast energy in the stimulus. Two forms of microbalanced stimuli are used, allowing us to test KDE shape identification performance with stimuli to which 1st-order motion detectors are blind. Finally, we examine stimuli in which moving textured tokens are camouflaged by a similarly textured background.

Method

Subjects. There were three subjects in this experiment. One was an author, and the other two were graduate students naive to the purposes of this experiment. All had normal or corrected-to-normal vision. There were slight differences in the conditions for each of the three subjects. These will be pointed out below.

White-on-gray dot stimuli. First, we briefly describe the stimuli that consist of bright dots moving on a gray background representing a variety of 3D shapes. This description will be somewhat abbreviated, since the same stimuli have been used in previous studies and more complete descriptions are available (Sperling et al., 1989). The other stimuli used in the present study result from simple image processing transformations applied to the white-on-gray dot stimuli.

Stimuli were based upon a fixed vocabulary of simple shapes consisting of bumps and concavities on a flat ground. The 3D shapes varied in the number, position, and 2D extent of these bumps and concavities. The process of generating the stimuli is illustrated in Fig. 1.

The first step in creating a stimulus involves the specification of a 3D surface. For a square area with sides of length s, a circle with diameter $0.9\,s$ is centered, and three fixed points, labeled 1, 2 and 3, are specified. For a given shape, one of two such sets of points is used (the upward-pointing triangle or the downward-pointing triangle, labeled u and d, respectively). The shape is specified as having a depth of zero outside of the circle. For each of the three identified points, the depth may be either $+0.5\,s$, 0.0, or $-0.5\,s$, which are labeled as $+$, 0, and $-$, respectively. The depth values for the rest of the figure were interpolated by using a standard cubic spline to connect the three interior points with the zero depth surround. Thus, there are 54 ways to designate a shape: u vs d, and for each of three interior points, $+$ vs 0 vs $-$. We designate a shape by denoting the triangle used, followed by the depth designations of the three points in the order shown in Fig. 1A. For example, $u-+0$

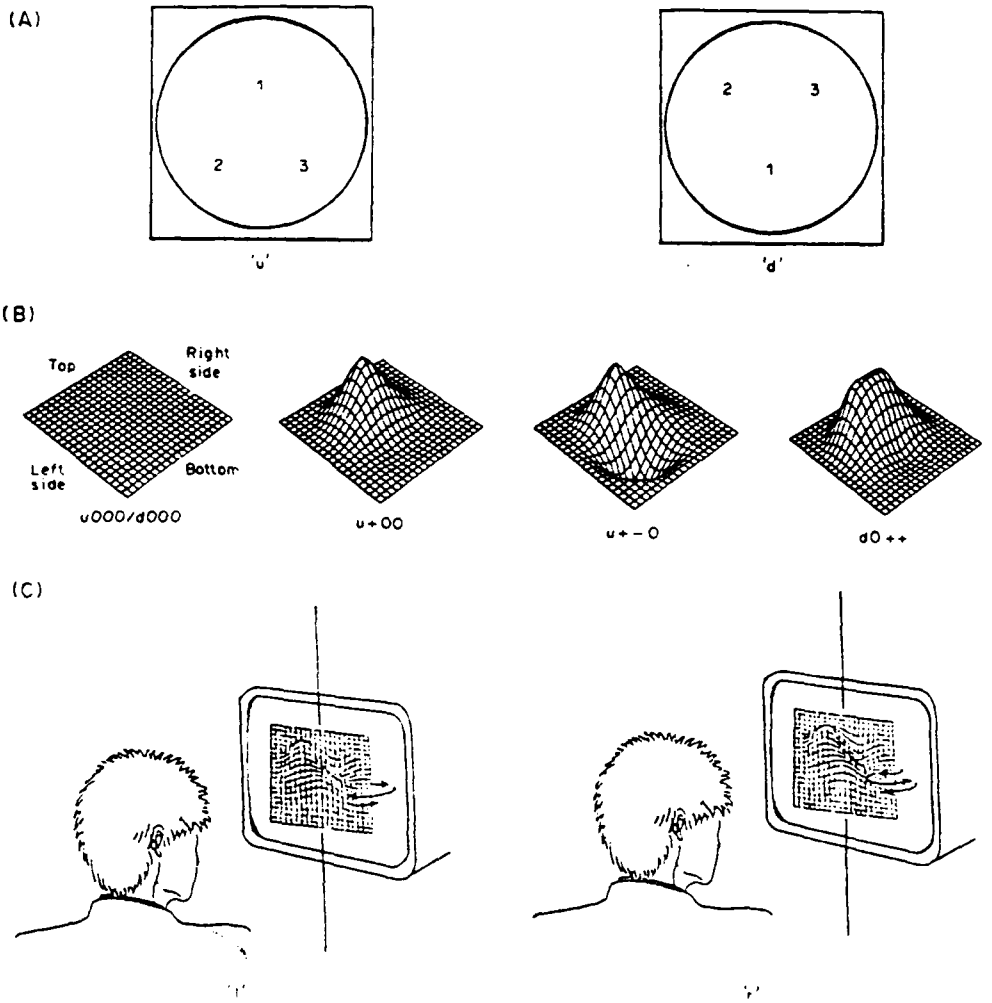


Fig. 1. Stimulus shapes, rotations, and their designations. (A) Shapes were constructed by choosing one of the two equilateral triangles represented here. Each point in the triangles was given a positive depth (i.e. toward the observer), zero depth, or negative depth, represented as +, 0 and −, respectively. A smooth shape splined these three points to zero depth values outside of the circle. A shape is designated by the choice of triangle (*u* or *d*), followed by the depth designations of the three points in the order given in the figure. (B) Some representative shapes generated by this procedure. All shapes consisted of a bump, concavity, or both, with a variation in position and extent of these areas. (C) Shapes were represented by a set of dots randomly painted on the surface of the shape, and wiggled about a vertical axis through the center of the display. The motion was a sinusoidal rotation that moved the object so as to face off to the observer's right, then his or her left, then back to face-forward (denoted *l*), or the reverse (denoted *r*).

is a shape with a bump in the upper-middle of the display, and a concavity in the lower-left (Fig. 1B). There are 53 distinct shapes, because *u*000 and *d*000 both denote a flat square.

Displays were generated by sprinkling dots randomly on the 3D surface generated by the spline, rotating that surface, and projecting the resulting dot positions onto the image plane using parallel perspective. A large number of dots are chosen uniformly over a 2D area somewhat larger than the *s* by *s* square, and each dot's depth is determined by the cubic spline interpolant (where the zero depth of the surround is continued outside the square). This collection of dots is rotated about a vertical axis that is at zero depth and centered in the display. The rotation angle $\theta(k)$ is a sinusoidal "wiggle": $\theta(k) = \pm 25 \sin(2\pi k/30)$ deg, where k is the frame number within the 30 frame display. Thus, the display either rotated 25 deg to the right, then reversed its direction until it faced 25 deg to the left, then reversed its direction until it was again facing forward (labeled *l*), or rotated in the opposite manner (labeled *r*, see Fig. 1C). The displays presented these 3D collections of dots in parallel perspective

as luminous dots (single pixels) on a darker background.

A stimulus name consists of the name of the shape followed by the type of rotation (e.g. $u+-0l$), resulting in 108 possible names. Using parallel perspective, there is a fundamental ambiguity with the KDE: reversing the depth values and rotation direction of a particular shape and rotation produces exactly the same display. In other words, a convexity rotating to the right produces exactly the same set of 2D dot motions as a concavity rotating to the left. Thus, $u+-0l$ and $u-+0r$ describe precisely the same display type. There is also no difference in display type among $u000l$, $u000r$, $d000l$ and $d000r$. This results in a total of 53 distinct display types.

These experiments used 54 white-on-gray dot displays, including two instantiations of the flat stimulus $u000$ (with different dot placements) and one instantiation of each other display type. Each set of dots was windowed to a display area of 182 × 182 pixels (corresponding to the $s \times s$ square), with dots presented as single luminous pixels.

When the dots on the surface of a shape move back and forth in the display, the local dot density changes as the steepness of the hills and valleys changes (with respect to the line of sight). In previous work (Sperling et al., 1989), we showed that this density cue is neither necessary nor sufficient for the perception of depth. However, it is a weak cue which one of three highly trained subjects was able to use for modest above-chance performance when it was presented in isolation. In other words, changing dot density is an artifactual cue to the task. As in previous experiments, we remove this cue by deleting or adding dots as needed throughout the display in order to keep local dot density constant. As a result of this manipulation, all displays had approx. 300 dots visible in the display window. The removal of the density cue results in a small amount of dot scintillation that neither lowers performance substantially nor appears to be useful as an artifactual cue (Sperling et al., 1989, 1990).

Other tokens. The 54 stimuli described so far consisted of luminous dots moving to and fro on a less luminous background. All other stimuli were based upon these displays. First, three conditions involved changes of the token that carried the motion. The moving dots were replaced with disks, patterned disks, or wires. We refer to the dot, wire, and disk conditions as *white-on-gray* stimuli, and the patterned disks as *pattern-on-gray*.

To create a disk stimulus, a dot stimulus is modified in the following way. Each luminous dot in the stimulus is replaced with a 6 × 6 pixel luminous diamond centered on the dot (Fig. 2b), which appears disk-like from the viewing distance used in the experiment. A sample image of white-on-gray disks is depicted in Fig. 2c, and is based on the white-on-gray dot stimulus frame shown in Fig. 2a.

The pattern-on-gray disk stimuli are generated in a similar fashion. The 6 × 6 diamond consists of 24 pixels which are a mixture of black and white (12 of each). These are displayed on an intermediate gray background. The diamond pattern and a sample stimulus frame are shown in Fig. 2d and e, respectively. Note that the diamond pattern has an equal number of black and white pixels in each row.

Other stimuli were based on "wires". Each dot was connected by a straight line (subject to the pixel sampling density) to all neighbors that were at a 2D distance no greater than 15.5 pixels (Fig. 2f). Note that a vector is drawn between two points based on their distance *in the image*, not on their simulated 3D distance. Since the lines were straight, when set in motion they objectively define a thickened surface with lines cutting through the interior of each bump and concavity. This may have yielded a perceived

Fig. 2 (*opposite*) Stimulus display generation for expt 1. (a) A single frame of a white-on-gray dots stimulus. All displays shown in this figure are based on this stimulus frame. (b) The diamond shape used to generate the disks from the dots. (c) A white-on-gray disks stimulus frame. (d) The patterned diamond for the pattern-on-gray condition. (e) A pattern-on-gray frame. (f) A white-on-gray wires frame. All pairs of dots in Fig. 2A were connected whose inter-point distance was less than 15.5 pixels. (g) A frame of dynamic-on-gray dots. In this condition each dot was painted black or white randomly and independently with probability of 0.5 for each color. (h) A frame of dynamic-on-gray disks. The same procedure as in (g) was applied to each pixel lying in each disk. (i) A frame of dynamic-on-gray wires. (j) A frame of dynamic-on-static disks. For both dynamic-on-static conditions (disks and wires), the tokens and the background consisted of random dot noise, and so the tokens cannot be discerned from a single static frame. (k) A frame of the pattern-on-static condition. This frame contains 300 copies of the pattern in (d) on a static noise background. The camouflage is quite effective. (l) An enlargement of the central portion of (k), with the patterned disks emphasized.

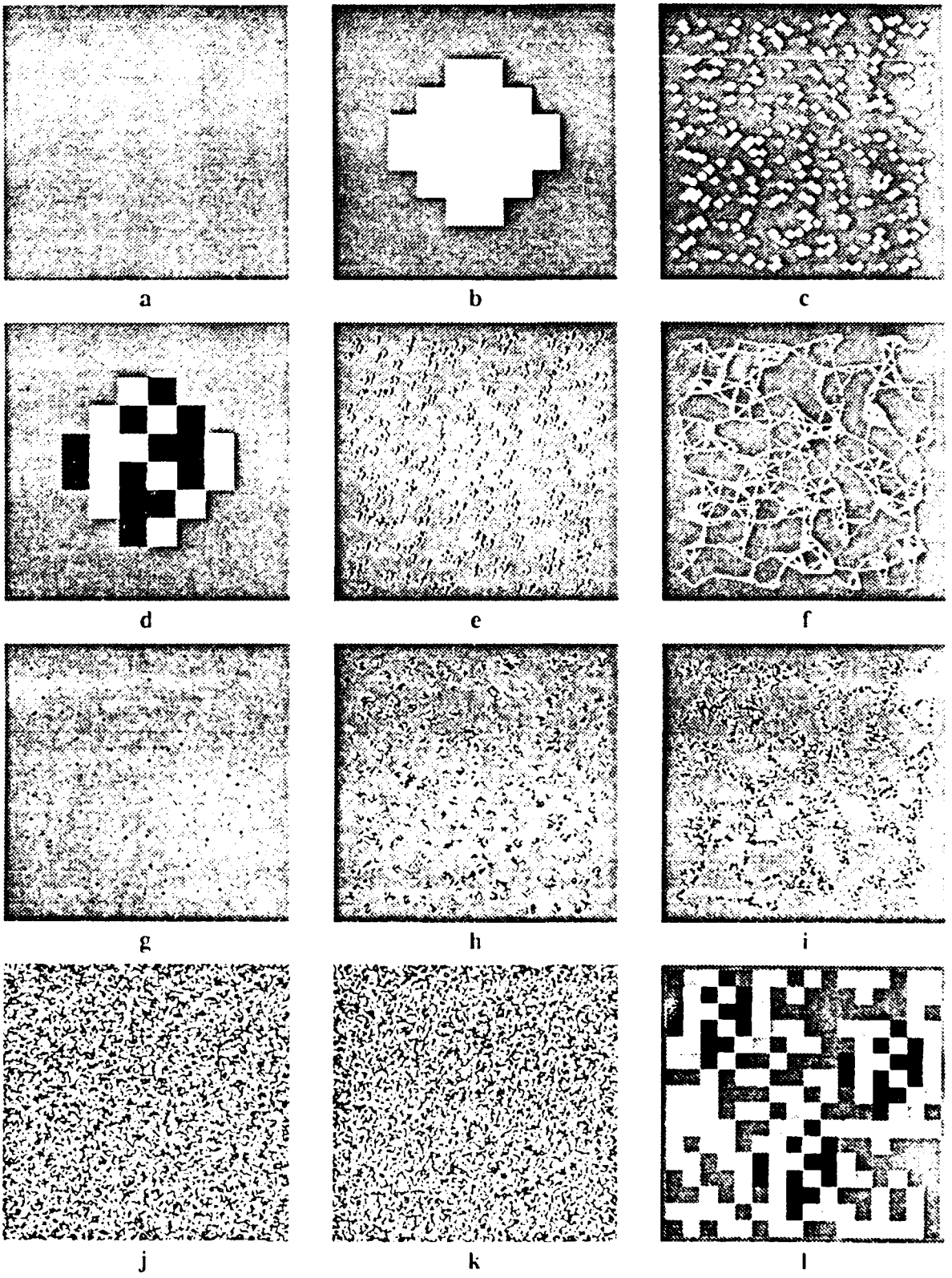


Fig. 2

(tesselated) surface having slightly less relative depth than the base surface. The choice of 15.5 pixels as the criterion for drawing a line was a compromise set in order to make sure that all stimulus dots became an endpoint to at least one line, and that no line was so long as to excessively cut through the simulated surface.

The white-on-gray disks and pattern-on-gray disks were based on the dot stimuli. The same exact instantiations were used in all three conditions. The nth frame of a given shape and rotation consisted of either dots, disks or patterned disks centered on the same set of image positions. For the wire stimuli, a new set of 54 instantiations was made.

Dynamic-on-gray. Three types of stimuli were used to explore the motion of patches of dynamic noise moving on a gray background. These stimuli are microbalanced, as we discussed in the previous section. These stimuli are derived from the dot, disk, and wire stimuli. To produce a dynamic-on-gray stimulus from a white-on-gray stimulus, simply change the luminance of each white pixel in each stimulus frame (i.e. the foreground or token pixels) to black randomly and independently with probability 0.5. Thus, foreground pixels undergo random contrast polarity alternation while background pixels are gray (i.e. have zero contrast). Sample frames are illustrated in Fig. 2g, h and i.

Dynamic-on-static. Two types of stimuli were used to explore the motion of patches of dynamic noise moving on a static noise background. This class of stimuli is also microbalanced (Chubb & Sperling, 1988b). We derive dynamic-on-static stimuli from the disk and wire stimuli. The foreground pixels consist of dynamic noise, just as in the previous dynamic-on-gray case. The background pixels consist of a static frame of patterned texture, where each pixel is randomly chosen to be either black or white with a probability of 0.5, just as the dynamic noise is. If a given pixel is a background position for two successive frames, then its color does not change. If that position is a foreground pixel in either or both frames, then there is a 50% chance that its color will change. A single frame of dynamic-on-static stimulus is simply a frame of random dot noise (Fig. 2j). The motion-carrying tokens are not discernible from a single frame. Rather, the areas of moving dynamic noise define the foreground tokens.

Contrast polarity alteration. Three stimulus conditions involved contrast polarity alternation. This stimulus manipulation was explored thoroughly for dot stimuli in the preceding paper (Dosher et al., 1989b). In this condition, the motion-carrying tokens alternate from white to black to white again on successive frames, all against a background of intermediate gray. Constrast polarity alternation was used with dots, disks, and wires, resulting in three polarity alternation conditions.

Pattern-on-static. The final condition involves pattern camouflage. This condition is derived from the pattern-on-gray stimuli. The gray background is replaced with a frame of static random dot noise. In other words, the patterned disk tokens move to and fro in front of a screen of static random dots, occluding it (and occasionally each other) as they pass by. A frame of this stimulus condition is pictured in Fig. 2k, and enlarged in Fig. 2l, where we have artificially highlighted the patterned disks for comparison to the pattern kernel shown in Fig. 2d. There are approx. 300 patterned disks in Fig. 2k. As you can see, the camouflage is quite effective. When the patterned disks move, as one might expect, they are easily visible (Julesz, 1971).

Display details. There are a total of 13 conditions (3 white-on-gray, 1 pattern-on-gray, 3 contrast polarity alternation, 3 dynamic-on-gray, 2 dynamic-on-static, and 1 pattern-on-static). There were 54 distinct displays for each of the 13 conditions. In all conditions, the displays are windowed to an area of 182 × 182 pixels. Displays were computed using the HIPS image processing software (Landy, Cohen & Sperling, 1984a, b), and displayed by an Adage RDS-3000 image display system.

Subjects MSL and JBL viewed these stimuli on a Conrac 7211C19 RGB color monitor. Only the green gun was used, and so stimuli appeared as bright green and black pixels (as dots, disks, lines or noise) on a green background of intermediate luminance. The stimuli subtended 3.7 × 4.2 deg. Stimuli were viewed monocularly through a dark viewing tunnel, using a circular aperture which was slightly larger than the stimuli.

Subject LJJ viewed the stimuli on a US Pixel PX15 black and white monitor with a P4-like phosphor. Here, stimuli subtended 2.9 × 2.9 deg, and appeared as white and black pixels on an intermediate gray background. Stimuli were viewed monocularly through a circular aperture in cardboard which approximately matched the hue of the displays, and

which had approximately the same luminance as the stimulus background.

Each stimulus consisted of 30 stimulus frames. These were presented at a 60 Hz frame rate. Each frame was repeated four times, resulting in an effective rate of 15 new stimulus frames per second. Each stimulus lasted 2 sec. A trial sequence consisted of a fixation spot, a blank interval, the 30 frame stimulus, and a blank. The fixation and blank lasted either for 1 sec each (subjects MSL and JBL), or 0.5 sec each (subject LJJ). The background luminance remained constant throughout the trial sequence. Subjects were free to use eye movements to actively explore the display. Stimuli were viewed from a distance of 1.6 m. After each stimulus display, subjects responded with the name of the shape and rotation direction using either a computer keyboard or response buttons.

Slightly different image luminances were used for each subject. The background luminance for subjects MSL, JBL and LJJ were 31.0, 40.0 and 45.0 cd/m^2 respectively. Since isolated luminous pixels were used, the appropriate unit of measurement is *extra* μcd/pixel for bright pixels, and *removed* μcd/pixel for dark pixels, all at a specified viewing distance (Sperling, 1971). Stimuli were calibrated so that extra μcd/pixel and removed μcd/pixel were equal. For subjects MSL, JBL and LJJ, these were 13.2, 19.2 and 15.7 μcd/pixel, respectively, at a viewing distance of 1.6 m. Contrasts were nominally 100%.

Procedure. There were 13 stimulus conditions. For each condition, there were 54 stimuli (two instantiations of the flat stimulus *u*000, and one instantiation of each of the 52 other possible distinct shape-rotation combinations). This resulted in 702 stimuli, each of which was viewed once by each subject. These 702 trials were viewed in random order in six blocks of 117 trials. On a given trial, a stimulus was shown, subjects keyed in their responses, and then feedback was provided so that we measured the best performance of which the subject was capable. Each block lasted approx. 1 hr. Subjects ran several practice sessions on the white-on-gray dots condition before data were collected. Given the mix of stimuli in a given condition, guessing base rates for the identification of shape and rotation direction were between 1.53 (for a strategy of random guessing) and 2.54 (for a strategy of always answering *u*000/, or one of its equivalents).

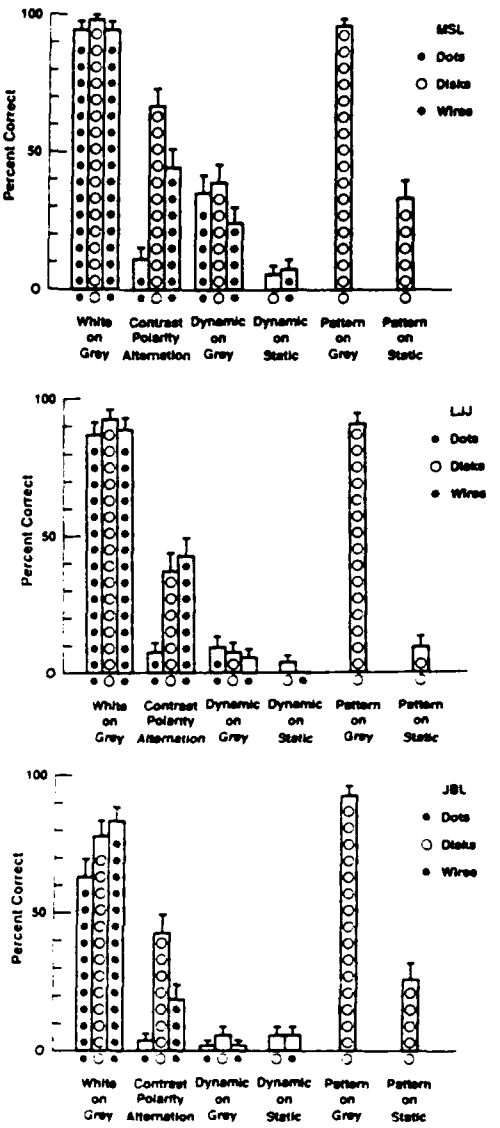


Fig. 3. Results of expt 1. Results are given for three subjects. Different symbols in the bars represent different tokens (large open dots for the disk and patterned disk tokens, small solid dots for the dot tokens, and asterisks for the wire tokens).

Results

The results for the three subjects are summarized in Fig. 3. Each performance measure given here is the percent correct over 54 trials. We discuss each class of stimulus condition in turn.

White-on-gray / Pattern-on-gray. As expected, the performance on the three white-on-gray and the one pattern-on-gray condition was uniformly high. The tokens provided excellent motion signals because they were moving rigid areas of high contrast. It did not particularly matter whether we used dots, as in our previous studies, wires, as in the early wire-frame KDE

work (Wallach & O'Connell, 1953), disks, or patterned disks. The disk and patterned disk stimuli provided very strong percepts of shape, although the disks did not undergo realistic foreshortening as they rotated. In fact, the dot stimuli gave the weakest percept of depth. These tokens had the least contrast energy (i.e. were the smallest), and hence were harder to detect. Subject JBL had the greatest difficulty in seeing these small dots, and his results show a slight drop in performance for the dot stimuli.

Dynamic-on-gray. The motion of a token filled with dynamic random dot noise moving on a gray background is microbalanced. In other words, 1st-order motion detectors are "blind" to this stimulus. The expected value of the output of such a detector is zero (across random realizations of the stimulus). Simple 2nd-order mechanisms (e.g. using rectification) serve to reveal the true motion.

The results for three subjects are somewhat different. For two subjects (LJJ and JBL), performance is always at or near chance (less than 10% correct in all cases), although for subject LJJ with the dynamic-on-gray dots the performance is significantly above chance ($P < 0.05$). On the other hand, for subject MSL, performance is always well above chance (24–39% correct identifications), but far less than his nearly perfect (94–98% correct) performance with white or pattern tokens on gray.*

The 1st-order motion mechanisms are clearly the most effective input to the KDE system, since eliminating motion detectable by 1st-order mechanisms reduces performance substantially for all subjects. The results for subject MSL suggest that 2nd-order motion mechanisms can also be used. On some trials, fragments of the microbalanced stimuli did appear 3D to this subject (one of the authors), especially in the foveally-viewed portion of the stimulus. To raise his performance level, he used sophisticated guessing strategies based on active eye movements and local measurements of motion or three-dimensionality in the fovea at a small number of locations of the display. But, these strategies only serve to bring performance up to mediocre levels in comparison with performance with rigid white-on-gray motion.

Dynamic-on-static. The dynamic-on-static manipulation also results in a micro-balanced stimulus. For the dynamic-on-static conditions, performance is at chance level for all three subjects, and for both wire disk tokens. As with the dynamic-on-gray conditions, the motion of the tokens is visible. It is not particularly difficult to detect the motion of an area of dynamic noise on a static noise background (Chubb & Sperling, 1988b). However, this sort of motion engenders no shape percept whatever under the conditions of our experiments.

Unlike dynamic-on-gray stimuli, dynamic-on-static stimuli are not revealed by contrast rectification. Detection of the motion of a region of flicker requires more elaborate 2nd-order mechanisms. Regions of flicker could first be detected by applying a linear temporal filter (such as differentiation), followed by rectification, and then by application of a 1st-order motion mechanism. Some such complex 2nd-order motion detector exists in the human visual system, since we are capable of seeing areas of flicker move, including in the displays of our experiment (at least with scrutiny). Yet, this 2nd-order motion detection system does not support the structure-from-motion computation for our dynamic-on-static stimuli.

Prazdny (1986) reached the opposite conclusion using dynamic-on-static displays representing simple wire objects rotating in a tumbling motion. Each object contained five wires, and subjects were required to identify the object among six alternative wire-frame objects.

*In order to test the range of luminances over which polarity alteration was effective, we ran a control experiment (using MSL and JBL as subjects), where a variety of white pixel luminances were used with a given black pixel luminance. We viewed a variety of dynamic-on-gray displays, varying the luminance values for the black and white pixels independently over a wide range. We also tested a variety of other luminance calibration procedures. Dynamic-on-gray stimuli are only micro-balanced if the contrast energy of the white pixels is the same as that of the black pixels. And, it is difficult to calibrate the luminance of individual pixels embedded in a complex display texture given that the desired pattern is first low-pass filtered by the CRT video amplifier, and then passes through the gun nonlinearity (see Mulligan & Stone, 1989, for a full discussion of this point). Thus, it was important to verify that our results were robust over a range of luminance values overlapping the calibrated equal contrast point.

To summarize, shape identification performance is consistent with the results of expt 1 for a reasonably wide range of white pixel luminances. Subject MSL consistently performs at moderate levels, and subject JBL consistently performs at or near chance. The luminance levels yielding poor shape identification performance are consistent with the levels that result in the weakest 3D percept, and are roughly consistent with the luminance levels that are balanced (black pixel decrement vs white pixel increment) for a variety of calibration displays. The performance levels for dynamic-on-gray stimuli in expt 1 do not result from a miscalibration of luminance levels.

The displays were 7 × 7 deg, and the wires were several pixels thick. Performance was quite high in the task for five subjects. Although we have some reservations about the experimental method employed by Prazdny, we have generated similar displays in our laboratory, and our dynamic-on-static wire-frame displays do yield a shape percept when displays are restricted to a small number of wires.

The most likely explanation of the difference between our results and those of Prazdny involves the difference in spatial resolution required by each task. Chubb and Sperling (1988a) have demonstrated that 2nd-order motion systems have less spatial resolution than the 1st-order mechanisms, and that their resolution drops precipitously with increases in retinal eccentricity. In our displays, motion was about a vertical axis using parallel perspective, and hence all motion was along the horizontal. There could be as many as 10 or 20 disks or wires in a given row of the image to resolve. Our displays did not yield a global percept of optic flow, but motion was perceived foveally with scrutiny. This is entirely consistent with Chubb and Sperling's observation. Prazdny did not give precise details about his stimuli, but it was clear that along a given motion path there were only two or three wires to resolve across his far larger display. Performance was so low in our dynamic-on-static conditions because too much spatial acuity was required of the 2nd-order system that detects the motion of flickering regions.

How useful for perception of shape is a display of dynamic noise figures moving on a static noise background? We have examined a large number of disk and (thick) wire displays in order to span the gap of spatial resolution between Prazdny's displays and our own. With our 3 × 3 deg display size, a shape percept can only be achieved by using a very small number of tokens (around 5–10). These displays consisted of rotating disk tokens. Cavanagh and Ramachandran (1988) suggest an alternative explanation of the difference between our results and those of Prazdny. They consider the crucial difference to be that the objects portrayed in the Prazdny displays were connected (one long wire figure), whereas our displays consisted of separate disk tokens. With our wire displays, almost no 3D percept was achieved for the dynamic-on-static condition. In addition, we were able to achieve a 3D percept with displays of a small number of dynamic-on-static disks. Thus, we feel that low spatial resolution in the 2nd-order motion system (rather than unconnected tokens) is the likely explanation for failure of KDE.

Contrast polarity alternation. Performance is quite poor for the contrast polarity-alternating dots as it was in the previous paper (Dosher et al., 1989b). For two subjects (JBL and LJJ) performance is at chance or insignificantly above chance. For subject MSL, performance is low (11% correct) but significantly above chance ($P < 0.05$). On the other hand, when the token is changed to disks or wires, performance rises substantially. Contrast polarity alternation is not as devastating a stimulus manipulation for disks and wires as it is for dots.

For 1st-order motion detection mechanisms such as the Reichardt detector, contrast polarity alternation causes the strongest responses to be in the wrong direction. Yet, the intended motion can be detected quite accurately if a 2nd-order detector is used that first applies a luminance nonlinearity followed by a Reichardt detector. The primary difference between the dots on the one hand, and the disks and wires on the other, is that the disks and wires have more pixels illuminated. In other words, they have more contrast energy, and in particular thay have more energy at lower spatial frequencies. Thus, the disk and wire stimuli should stimulate both the 1st- and 2nd-order motion detection systems more strongly, resulting in stronger incorrect direction information from the 1st-order system as a whole, but also stronger information from the 2nd-order system, and stronger directional information in those selected 1st-order frequency bands which signal the correct direction.

It is interesting to note that a large number of the errors made by observers with polarity-alternating stimuli were errors in the direction of rotation *only*, with the shape specified correctly. For example, for a stimulus which had as correct answers either $u+-0l$ or $u-+0r$, the subject incorrectly responded with $u+-0r$ or $u-+0l$, rather than with any of the 104 other possible incorrect responses. This effect was largest for the disk tokens. In a separate control experiment, for contrast polarity-alternating disk stimuli, 39% of the errors made by subject MSL were only an error in the specification of direction, compared to 1.4% direction errors for the dynamic-on-gray conditions. For subject JBL, the corresponding values were 48% and 5.6%. For the polarity-alternating disks, on

trials when subject MSL correctly identified the shape, there was a 33% chance that he would misidentify the direction of rotation (for JBL: 29.3%). We believe that accurate shape identification in this condition primarily reflects responses constructed from selected 1st-order information. One strategy was simply to specify *the opposite rotation direction to that which* was perceived! The displays did, however, occasionally appear to be 3D with the correct direction of motion (at certain times during the rotation, or close to the location to which the eyes were directed), indicating a residual 2nd-order motion input to the KDE system. The fact that these displays only appeared foveally to be rotating in the correct direction, and then only using the larger tokens, is consistent with a 2nd-order motion detection system with low contrast sensitivity and low spatial resolution (as has been demonstrated by Chubb & Sperling, 1988b), and more sensitive in the fovea (Chubb & Sperling, 1988a). In summary, we have some indication that 2nd-order motion detection mechanisms can be used to derive 3D structure, but they are far less robust and have poorer spatial resolution than 1st-order motion mechanisms.

Pattern-on-static. For all three subjects performance with pattern-on-static displays is quite poor (9, 26 and 33% corrrect), although it is significantly above chance levels in all cases ($P < 0.05$). This poor performance results from a mismatch of resolution and temporal sampling. The patterned disks are quite detailed/high frequency. The disks are 6 pixels in diameter, and can move as far as 8.3 pixels in one frame. This speed is only achieved by disks at the top of a peak when in the middle of the display (i.e. near frame numbers 0, 15 and 29), but many disks are moving 3–5 pixels per frame. High frequency spatial filters which are required to identify the disks must correlate across frames with filters that are far more than 90 deg away in the phase of their peak spatial frequency. A typical 1st-order detector will not compare spatial regions that far apart in order to avoid spatio-temporal aliasing (van Santen & Sperling, 1984). Thus, the clearest motion signals are coming from the slower areas in the display, which are the least useful for discriminating the shapes. We have examined pattern-on-static displays with finer temporal sampling (60 new frames per sec, as opposed to 4 repaints of 15 new frames per sec used in the experiment), and they give a strong impression of three-dimensionality. Thus, poor performance in the task resulted from undersampling in time of the stimuli, which interferes with 1st-order (and some 2nd-order) motion mechanisms, and good KDE can result from the motion of tokens which are camouflaged when at rest.

We have also examined dynamic-on-static displays with finer temporal sampling (60 new frames per sec). These displays yield no impression of three-dimensionality. The poor results for dynamic-on-static displays do not result from insufficient sampling in time. Also, since finely sampled pattern-on-static displays do appear 3D, poor performance with dynamic-on-static-displays does not result from the camouflage of the tokens when at rest. Rather, dynamic-on-static displays yield no effective KDE because of the low resolution of the 2nd-order system required to analyze the motion.

EXPERIMENT 2. TWO-FRAME KDE

The first experiment shows that accurate performance in shape identification is dependent upon a global (primarily 1st-order) optic flow. If a stimulus manipulation makes that optic flow noisy or otherwise interferes with the optic flow computation, there is little or no KDE. This occurs even though foveal scrutiny does reveal the motion in these displays.

If the percept of surface shape depends upon a global optic flow, then we should be able to get reasonable shape identification performance from any stimulus that results in a strong percept of optic flow. In particular, the extended (2 sec) viewing conditions of expt 1 should not be necessary. Two frames are obviously the minimum number of frames that can yield a percept of motion, and two frames should suffice. In the second experiment, we investigate the accuracy of performance in the shape identification task for two-frame displays.

Method

Subjects. There were two subjects in this experiment. One was an author, and the other was a graduate student naive to the purposes of this experiment. Both had normal or corrected-to-normal vision. There were slight differences in the conditions for each of the two subjects. These will be pointed out below.

Stimuli and apparatus. The stimuli were similar to the white-on-gray dot stimuli from expt 1. Stimuli were generated from the same set of 3D

shapes, using the same dot densities, and projected in the same way. The local dot density was kept constant using the same scintillation procedure. New stimuli were computed, two of the flat shape, and one of each of the other 52 shapes, resulting in 54 displays.

Each display consisted of 11 frames, rotating from 20 deg left to 20 deg right in increments of 4 deg per frame. The middle frame (number 6) was face-forward, as was the first frame of each display in expt 1. Two-frame stimuli consisted of a presentation of the middle frame followed by one of the other 10 display frames. This resulted in either a leftward or rightward rotation of 4–20 deg between the two frames of the display. A single trial display consisted of 0.5 sec of a cue spot, 0.5 sec blank, the first frame, an inter-stimulus blank interval (or ISI), the second frame, and a blank. Each stimulus frame was repainted four times at 60 Hz, for a total duration of 67 msec. We define the ISI to be the time interval between the onset of the last painting of the first stimulus frame and the onset of the first painting of the second stimulus frame. For example, when no blank frames were used, the ISI was 16.7 msec. Displays were 182×182 pixels, and were presented using the same apparatus and viewing conditions as for subject LJJ in expt 1. The background luminances for subjects MSL and LJJ were 15.6 cd/m^2 and 5.0 cd/m^2, respectively. The corresponding dot luminosities were 26.8 and 15.7 extra μcd/dot, respectively. Nominal contrasts were huge (i.e. nominal Weber contrasts of 500% or more).

Procedure. The task was shape and rotation identification. Subjects keyed their responses using response buttons, and received feedback on the display after their response. Three groups of trials were run. In the first, the ISI was 16.7 msec, and rotation angle between frames was varied from 4 to 20 deg. Since the second frame could be chosen from either the frames preceding or succeeding the middle frame (rotation to the left or right), this resulted in 540 possible stimuli (54 displays, 2 directions, 5 rotation angles). These were run in random order in 4 blocks of 135 trials. In the second group of trials, rotation was kept constant at 4 deg. ISI ranged from 16.7 to 83.3 msec. This again resulted in 540 trials presented in random order in 4 blocks of 135 trials. In the third group

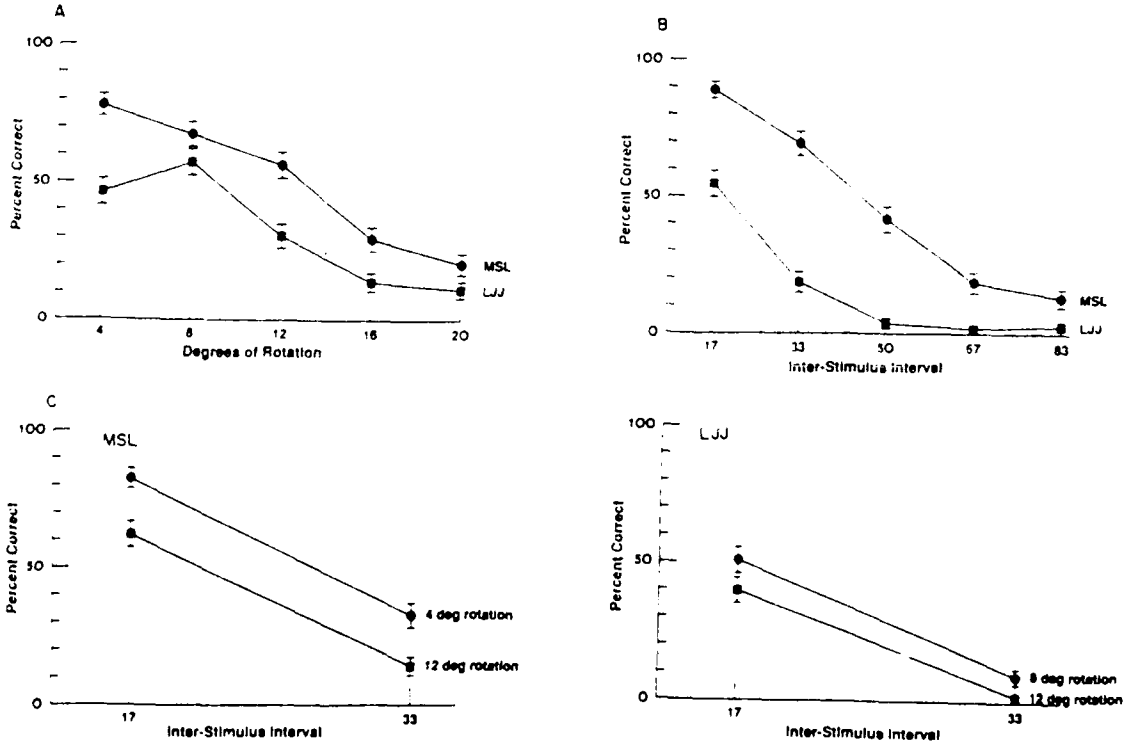


Fig. 4. Results of expt 2. Data for two subjects are shown. Error bars indicate ± 1 SEM. (A) Shape-and-rotation identification accuracy as a function of the angle of rotation between the two frames. ISI was 16.7 msec. (B) Shape-and-rotation identification accuracy as a function of the duration of a blank inter-stimulus interval (ISI). Rotation angle was 4 deg. (C) The two manipulations used in the same experiment. Note the lack of interaction.

of trials, both rotation angle and ISI were varied. The ISIs were either 16.7 or 33.3 msec. For subject MSL, the rotation angles were either 4 or 12 deg. For LJJ, they were either 8 or 12 deg. These four conditions (two rotation angles by two ISIs) resulted in 432 trials which were presented in random order in 4 blocks of 108 trials.

Results

The results are shown in Fig. 4. Each data point is the percent correct over 108 trials. As is evident from the figure, shape identification can be quite high for these minimal motion displays (for similar observations using different experimental methodology, see Braunstein, Hoffman, Shapiro, Andersen & Bennett, 1987; Lappin, Doner & Kottas, 1980; Mather, 1989; and Petersik, 1980). For an ISI of 16.7 msec (Fig. 4A), this entire sequence lasted only 133 msec. Yet, performance was as high as 54.6% for subject LJJ, and 88.9% for subject MSL (62.8% and 94.2% of their white-on-gray dots performance in expt 1, respectively). Two frames of moving dots are sufficient for accurate, although not perfect
performance in this shape identification task. Since these experiments were first reported (Landy, Sperling, Dosher & Perkins, 1987a; Landy, Sperling, Perkins & Dosher, 1987b), Todd (1988) has also shown above-chance KDE performance for two-frame stimuli, although in his paradigm the two frames are repeated several times before a response is made.

Rotation angle and fixation. Performance as a function of rotation angle between the two frames is given in Fig. 4A. Performance decreases with increasing angle of rotation for subject MSL. For subject LJJ, performance reaches a peak at 8 deg, and decreases for smaller and larger rotations. The decrease in performance with larger rotation angles is to be expected, since the correspondence problem becomes increasingly difficult as dots move farther from their initial positions. One might also expect performance to drop as rotation angle decreases to zero. At extremely small rotation angles, the remaining motion would fall below threshold. In our displays, the drop with small rotation angles might be expected to occur even sooner as the small motions in the display became corrupted by poor spatial sampling (inter-pixel distance was approx. 1 min arc). This drop was only seen in the data of LJJ, and presumably would be seen in those of MSL if he had been tested using smaller rotations.

In a previous paper (Dosher et al., 1989b), we found that adding a blank interval between successive frames of a 30 frame KDE stimulus reduced shape identification to near chance performance. This was explained by reduction of power in the stimulus to the 1st-order system. This effect is also seen here, where performance decreases monotonically with increasing ISI (Fig. 4B). Subject LJJ performs at chance levels with a 50 msec or greater ISI, while subject MSL is still slightly above chance performance with an 83.3 msec ISI.

Time and distance. In the previous two groups of trials, there was a confounding between the stimulus manipulation (rotation angle or ISI) and dot velocity. Greater rotation angles at a fixed (16.7 msec) ISI produced greater velocities. Similarly, greater ISIs at a fixed 4 deg rotation angle resulted in smaller velocities. If performance were simply a function of velocity, then rotation angle and ISI should trade off. In Fig. 4C we present the results of varying both ISI and rotation angle factorially. We used a different set of rotations for subject LJJ than MSL based on the results in Fig. 4A, so that for both subjects the performance was expected to decrease with increasing rotation angles. As can be seen in the figure, the two variables do not trade off as would be expected if performance were only a function of velocity, or rotation speed. Increasing rotation angle increases the difficulty of the correspondence problem. Increasing ISI causes increasing problems for the motion detection system. Both manipulations degrade performance in an additive fashion. This observation contradicts Korte's (1915) 3rd law of apparent motion perception, which states that an increase in ISI must be counteracted by an increase in distance traveled for strong apparent motion. In Fig. 4C, Korte's law predicts a cross-over interaction, which is strongly disconfirmed. However, Burt and Sperling (1981) show that time and distance have independent additive effects on the strength of the apparent motion of dot stimuli, which agrees with the present results.

KDE from optic flow. Accurate KDE performance requires a global optic flow. When that optic flow is produced by a minimal motion stimulus—a two-frame display—the shape percept may be fragile and easily degraded by a variety of stimulus manipulations. The stimuli are quite brief in this paradigm and, by subject

reports, appear as a collection of dots moving at various speeds, i.e. "look like" an optic flow. On some trials, only patches of planar motion are perceived, and the shape response is generated cognitively. On other trials, a 3D surface is perceived. On some trials the optic flow is perceived and so is the shape, but the shape percept is only "felt" after the display is over. As we discussed extensively in our first article on the shape identification task (Sperling et al., 1989), KDE is inextricably tied with the percept of an optic flow. It can be very difficult to differentiate empirically between a judgment based on a 3D percept and performance based on an alternative strategy (computationally equivalent to that required for KDE) using a remembered set of 2D velocities.

Reasonably accurate performance on the shape-and-rotation identification task results from only two frames of 300 points. In the computer vision literature, there have been several studies of the structure-from-motion problem resulting in theorems of the following form: "*m* views of *n* points under the following restrictions of the motion path suffice to determine the 3D structure up to a reflection" (Bennett & Hoffman, 1985; Hoffman & Bennett, 1985; Hoffman & Flinchbaugh, 1982; Ullman, 1979). It has been suggested that these minimal conditions for structure from motion also govern human perception (Braunstein et al., 1987; Petersik, 1987). The particular models just mentioned do not have any prediction concerning performance in the 300 points 2 views situation used here. An exception is a recent paper by Bennett, Hoffman, Nicola and Prakash (1989), where it is shown that there is a one parameter family of possible interpretations for two frames of four or more points. This family is parameterized by the slant of the axis of rotation (as in the "isokinescopic displays" described by Adelson, 1985), and the paper does not deal explicitly with rotation axes in the image plane, as used here. On the other hand, models that compute 3D structure based only upon a single velocity field do allow for this performance (Longuet-Higgins & Prazdny, 1980; Koenderink & van Doorn, 1986). We take our experimental results as evidence for optic flow-based methods for the KDE, as opposed to models requiring three or more views. In particular, our results strongly rule out models that require measurement of acceleration in addition to velocity (e.g. Hoffman, 1982).

Structure-from-motion computation may improve its 3D representation with additional information (e.g. with additional frames, Grzywacz, Hildreth, Inada & Adelson, 1988; Hildreth & Grzywacz, 1986; Landy, 1987; Ullman, 1984). The shape in our two-frame displays does not always appear to have the depth extent that results from the 30 frame displays of expt 1, and two-frame performance is reduced relative to 30-frame performance. The shape identification task can be solved by knowing only the sign of depth and direction of motion in each spatial location (up to a reflection), without accurately estimating either velocity or the amount of depth.

DISCUSSION

Two experiments investigated the type of motion detection mechanism used as an input to the structure-from-motion system. Performance in the shape-and-rotation identification task was accurate regardless of the token used to carry the motion, as long as that token was presented with constant contrast polarity (the white-on-gray and pattern-on-gray conditions). The performance decrements seen with contrast polarity alternation and the two microbalanced conditions add further evidence to the conclusion of Dosher et al. (1989b) that 1st-order motion detectors are the primary substrate for the computation of shape. In addition, there are indications of an input to the shape computation from 2nd-order motion mechanisms, which is weak, low in spatial resolution, and concentrated at the fovea. 2nd-order mechanisms that require temporal filtering (i.e. detection of flicker) prior to a point nonlinearity were useless here because of the spatial resolution required by our stimuli. These sorts of detectors would only be useful for KDE displays involving a small number of moving features, rather than the densely sampled optic flows required for the determination of precise shapes of curved surfaces from motion cues. The results from the two-frame experiments reinforced these conclusions. They also demonstrated that detection of instantaneous velocity is sufficient for KDE; acceleration is not required, nor are more than two views.

Acknowledgements—The work described in this paper was supported primarily by a grant from the Office of Naval Research, grant N00014-85-K-0077, and partly by USAF Life Science Directorate, grants 85-0364, 88-0140, and NSF grant IST-8418867. We would like to thank Charles Chubb

for his helpful comments, and Robert Picardi for technical assistance. Portions of this work have been presented at the annual meetings of the Association for Research on Vision and Ophthalmology, Sarasota, Florida (Landy et al., 1987a) and the Optical Society of America, Rochester, New York (Landy et al., 1987b).

REFERENCES

Adelson, E. H. (1985). Rigid objects appear highly non-rigid. *Investigative Ophthalmology and Visual Science* (Suppl.), *26*, 56.

Adelson, E. H. & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A*, *2*, 284–299.

Anstis, S. M. & Rogers, B. J. (1975). Illusory reversal of depth and movement during changes of contrast. *Vision Research*, *15*, 957–961.

Bennett, B. M. & Hoffman, D. D. (1985). The computation of structure from fixed-axis motion: Nonrigid structures. *Biological Cybernetics*, *51*, 293–300.

Bennett, B. M., Hoffman, D. D., Nicola, J. E. & Prakash, C. (1989). Structure from two orthographic views of rigid motion. *Journal of the Optical Society of America A*, *6*, 1052–1069.

Braunstein, M. L., Hoffman D. D., Shapiro, L. R., Andersen, G. J. & Bennett, B. M. (1987). Minimum points and views for the recovery of three-dimensional structure. *Journal of Experimental Psychology: Human Perception and Performance*, *13*, 335–343.

Burr, D. C., Ross, J. & Morrone, M. C. (1986). Seeing objects in motion. *Proceedings of the Royal Society of London, B*, *227*, 249–265.

Burt, P. & Sperling, G. (1981). Time, distance, and feature trade-offs in visual apparent motion. *Psychological Review*, *88*, 171–195.

Cavanagh, P. & Ramachandran, V. S. (1988). Structure from motion with equiluminous stimuli. Paper presented to the *Annual Meeting of the Canadian Psychological Association*, Montreal, June.

Chubb, C. & Sperling, G. (1988a). Processing stages in non-Fourier motion perception. *Investigative Ophthalmology and Visual Science (Suppl.)*, *29*, 266.

Chubb, C. & Sperling, G. (1988b). Drift-balanced random stimuli: A general basis for studying non-Fourier motion perception. *Journal of the Optical Society of America A*, *5*, 1986–2007.

Chubb, C. & Sperling, G. (1989a). Two motion perception mechanisms revealed through distance-driven reversal of apparent motion. *Proceedings of the National Academy of Sciences, U.S.A.*, *86*, 2985–2989.

Chubb, C. & Sperling, G. (1989b). Second-order motion perception: Space/time separable mechanisms. *Proceedings: Workshop on visual motion* (pp. 126–138). Washington, D.C.: IEEE Computer Society Press.

Dosher, B. A., Landy, M. S. & Sperling, G. (1989a). Ratings of kinetic depth in multi-dot displays. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 816–825.

Dosher, B. A., Landy, M. S. & Sperling, G. (1989b). The kinetic depth effect and optic flow—I. 3D shape from Fourier motion. *Vision Research*, *29*, 1789–1813.

Grzywacz, N. M., Hildreth, E. C., Inada, V. K. & Adelson, E. H. (1988). The temporal integration of 3-D structure from motion: A computational and psychophysical study. In von Seelen, W., Shaw, G. & Leinhos, U. M. (Eds.), *Organization of neural networks*. New York: VCH.

Heeger, G. J. (1987). Model for the extraction of image flow. *Journal of the Optical Society of America A*, *4*, 1455–1471.

Hildreth, E. C. & Grzywacz, N. M. (1986). The incremental recovery of structure from motion: Position vs velocity based formulations. *Proceedings of the workshop on motion: Representation and analysis*. IEEE Computer Society no. 696, Charleston, South Carolina, 7–9 May.

Hoffman, D. D. (1982). Inferring local surface orientation from motion fields. *Journal of the Optical Society of America 72*, 888–892.

Hoffman, D. D. & Bennett, B. M. (1985). Inferring the relative three-dimensional positions of two moving points. *Journal of the Optical Society of America A*, *2*, 350–353.

Hoffman D. D. & Flinchbaugh, B. E. (1982). The interpretation of biological motion. *Biological Cybernetics*, *42*, 195–204.

Julesz, B. (1971). *Foundations of cyclopean perception*. Chicago, IL: The University of Chicago Press.

Koenderink, J. J. & van Doorn, A. J. (1986). Depth and shape from differential perspective in the presence of bending deformations. *Journal of the Optical Society of America A*, *3*, 242–249.

Korte, A. (1915). Kinematoskopische Untersuchungen. *Zeitschrift für Psychologie*, *72*, 193–206.

Landy, M. S. (1987). A parallel model of the kinetic depth effect using local computations. *Journal of the Optical Society of America A*, *4*, 864–876.

Landy, M. S., Cohen, Y. & Sperling, G. (1984a). HIPS: A Unix-based image processing system. *Computer Vision, Graphics and Image Processing*, *25*, 331–347.

Landy, M. S., Cohen, Y. & Sperling, G. (1984b). HIPS: Image processing under UNIX—Software and applications. *Behavior Research Methods, Instruments and Computers*, *16*, 199–216.

Landy, M. S., Sperling, G., Dosher, B. A. & Perkins, M. E. (1987a). Structure from what kinds of motion? *Investigative Ophthalmology and Visual Science (Suppl.)*, *28*, 233.

Landy, M. S., Sperling, G., Perkins, M. E. & Dosher, B. A. (1987b). Perception of complex shape from optic flow. *Journal of the Optical Society of America A*, *4*, 108.

Lappin, J. S., Doner, J. F. & Kottas, B. L. (1980). Minimal conditions for the visual detection of structure and motion in three dimensions. *Science*, *209*, 717–719.

Lelkens, A. M. M. & Koenderink, J. J. (1984). Illusory motion in visual display. *Vision Research*, *24*, 1083–1090.

Longuet-Higgins, H. C. & Prazdny, K. (1980). The interpretation of a moving retinal image. *Proceedings of the Royal Society of London B*, *208*, 385–397.

Marr, D. & Ullman, S. (1981). Directional selectivity and its use in early visual processing. *Proceedings of the Royal Society of London B*, *211*, 151–180.

Mather, G. (1989). Early motion processes and the kinetic depth effect. *The Quarterly Journal of Experimental Psychology*, *41A*, 183–198.

Mulligan, J. B. & Stone, L. S. (1989). Halftoning method for the generation of motion stimuli. *Journal of the Optical Society of America A*, *6*, 1217–1227.

Petersik, J. T. (1980). The effects of spatial and temporal factors on the perception of stoboscopic rotation simulations. *Perception*, *9*, 271–283.

Petersik, J. T. (1987). Recovery of structure from motion: Implications for a performance theory based on the structure-from-motion theorem. *Perception and Psychophysics, 42*, 355–364.

Prazdny, K. (1986). Three-dimensional structure from long-range apparent motion. *Perception, 15*, 619–625.

Ramachandran, V. S., Rao, V. M. & Vidyasagar, T. R. (1973). Apparent movement with subjective contours. *Vision Research, 13*, 1399–1401.

Reichardt, W. (1957). Autokorrelationsauswertung als Funktionsprinzip des Zentralnervensystems. *Zeitschrift Naturforschung B, 12*, 447–457.

van Santen, J. P. H. & Sperling, G. (1984). Temporal covariance model of human motion perception. *Journal of the Optical Society of America A, 1*, 451–473.

van Santen, J. P. H. & Sperling, G. (1985). Elaborated Reichardt detectors. *Journal of the Optical Society of America A, 2*, 300–321.

Sperling, G. (1971). The description and luminous calibration of cathode ray oscilloscope visual displays. *Behavior Research Methods and Instruments, 3*, 148–151.

Sperling, G. (1976). Movement perception in computer-driven visual displays. *Behavior Research Methods and Instrumentation, 8*, 144–151.

Sperling, G., Landy, M. S., Dosher, B. A. & Perkins, M. E. (1989). The kinetic depth effect and identification of shape. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 826–840.

Sperling, G., Dosher, B. A. & Landy, M. S. (1990). How to study the kinetic depth effect experimentally. *Journal of Experimental Psychology: Human Perception and Performance 16*, 445–450.

Todd, J. T. (1988). Perceived 3D structure from 2-frame apparent motion. *Investigative Ophthalmology and Visual Science (Suppl.), 29*, 265.

Ullman, S. (1979). *The interpretation of visual motion*. Cambridge, MA: MIT Press.

Ullman, S. (1984). Maximizing rigidity: The incremental recovery of 3-D structure from rigid and non-rigid motion. *Perception, 13*, 255–274.

Wallach, H. & O'Connell, D. N. (1953). The kinetic depth effect. *Journal of Experimental Psychology, 45*, 205–217.

Watson, A. B. & Ahumada, A. J. Jr (1985). Model of human visual-motion sensing. *Journal of the Optical Society of America A, 1*, 322–342.

Vision Res. Vol. 31, No. 7/8, pp. 1399–1415, 1991
Printed in Great Britain.

0042-6989/91 $3.00 + 0.00

OBJECT SPATIAL FREQUENCIES, RETINAL SPATIAL FREQUENCIES, NOISE, AND THE EFFICIENCY OF LETTER DISCRIMINATION

DAVID H. PARISH and GEORGE SPERLING*

Human Information Processing Laboratory, Department of Psychology and Center for Neural Sciences, New York University, NY 10003, U.S.A.

(Received 7 July 1988; in revised form 2 June 1990)

Abstract—To determine which spatial frequencies are most effective for letter identification, and whether this is because letters are objectively more discriminable in these frequency bands or because we can utilize the information more efficiently, we studied the 26 upper-case letters of English. Six two-octave wide filters were used to produce spatially filtered letters with 2D-mean frequencies ranging from 0.4 to 20 cycles per letter height. Subjects attempted to identify filtered letters in the presence of identically filtered, added Gaussian noise. The percent of correct letter identifications vs s/n (the root-mean-square ratio of signal to noise power) was determined for each band at four viewing distances ranging over 32:1. Object spatial frequency band and s/n determine *presence of information* in the stimulus; viewing distance determines retinal spatial frequency, and affects only *ability to utilize*. Viewing distance had no effect upon letter discriminability: object spatial frequency, not retinal spatial frequency, determined discriminability. To determine discrimination efficiency, we compared human discrimination to an ideal discriminator. For our two-octave wide bands, s/n performance of humans and of the ideal detector improved with frequency mainly because linear bandwidth increased as a function of frequency. Relative to the ideal detector, human efficiency was 0 in the lowest frequency bands, reached a maximum of 0.42 at 1.5 cycles per object and dropped to about 0.104 in the highest band. Thus, our subjects best extract upper-case letter information from spatial frequencies of 1.5 cycles per object height, and they can extract it with equal efficiency over a 32:1 range of retinal frequencies, from 0.074 to more than 2.3 cycles per degree of visual angle.

Spatial filtering Scale invariance Psychophysics Contrast sensitivity Acuity

INTRODUCTION

Characterizing objects

When we view objects, what range of spatial frequencies is critical for recognition, and how is our visual system adapted to perceive these frequencies? Ginsburg (1978, 1980) was among the first to investigate this problem by means of spatial bandpass filtered images of faces and lowpass filtered images of letters. He noted the lowest frequency band for faces and the cutoff frequency for letters at which the images seemed to him to be clearly recognizable. The cutoff frequency for letters was 1–2 cycles per letter width; faces were best recognized in a band centered at 4 cycles per face width. He also proposed that the perception of geometric visual illusions, such as the Mueller–Lyer and Poggendorf, was mediated by low spatial frequencies (Ginsberg, 1971, 1978; Ginsberg & Evans, 1979).

An issue that is related to the lowest frequency band that suffices for recognition is the encoding economy of a band. For a filter with a bandwidth that is proportional to frequency (e.g. a two-octave-wide filter), the lower the frequency, the smaller the number of frequency components needed to encode the filtered image of a constant object. Combining these two notions, Ginsburg concluded that objects were best, or most efficiently, characterized by the lowest band of spatial frequencies that sufficed to discriminate them. Ginsburg (1980) went on to suggest that higher spatial frequencies were redundant for certain tasks, such as face or letter recognition.

Several investigators were quick to point out that objects can be well discriminated in various spatial frequency bands. Fiorentini, Maffei and Sandini (1983) observed that faces were well recognized in either high or in lowpass filtered bands. Norman and Erlich (1987) observed that high spatial frequencies were essential for discrimination between toy tanks in photographs.

*To whom reprint requests should be addressed.

With respect to geometric illusions, both Janez (1984) and Carlson, Moeller and Anderson (1984) observed that the geometric illusions could be perceived for images that had been highpass filtered so that they contained no low spatial frequencies. This suggests that low and high spatial frequency bands may carry equivalently useful information for higher visual processes.

Characterizing the visual system

In the studies cited above, the discussion of spatial filtering focuses on *object* spatial frequencies, that is, frequencies that are defined in terms of some dimension of the object they describe (cycles per object). Most psychophysical research with spatial frequency bands has focused on *retinal* spatial frequencies, that is, frequencies defined in terms of retinal coordinates. For example, the spatial contrast sensitivity function (Davidson, 1968; Campbell & Robson, 1968) describes the threshold sensitivity of the visual system to sine wave gratings as a function of their *retinal* spatial frequency. Visual system sensitivity is greatest at 3–10 cycles per degree of visual angle (c/deg). How does visual system sensitivity relate to object spatial frequencies?

Unconfounding retinal and object spatial frequencies

Retinal spatial frequency and object spatial frequency can be varied independently to determine whether certain object frequencies are best perceived at particular retinal frequencies. Object frequency is manipulated by varying the frequency band of bandpass filtered images; retinal frequency is manipulated by varying the viewing distance.

The cutoff *object* spatial frequency of lowpass filters and the observer's viewing distance were varied independently by Legge, Pelli, Rubin and Schleske (1985) who studied reading rate of filtered text at viewing distances over a 133:1 range. Over about a 6:1 middle range of distances, reading rate was perfectly constant, and it was approximately constant over a 30:1 range. At the longest viewing distances, there was a sharp performance decrease (as the letters became indiscriminably small). At the shortest viewing distance, performance decreased slightly, perhaps due to large eye movements that the subjects would have to execute to bring relevant material towards their lines of sight, and to the impossibility of peripherally previewing new text.

While viewing distance changed the overall level of performance in Legge et al., the cutoff *object* frequency of their low-pass filters at which performance asymptoted did not change. From this study, we learn that reading rate can be quite independent of retinal frequency over a fairly wide range, and that dependence on critical object frequency does not depend on viewing distance. Because the authors measured reading rate only in lowpass filtered images, we cannot infer reading performance in higher spatial frequency bands from their data.

Unconfounding object statistics and visual system properties

Human visual performance is the result of the combined effects of the objectively available information in the stimulus, and the ability of humans to utilize the information. In studying visual performance with differently filtered images, it it critical to separate availability from ability to utilize. For example, narrow-band images can be completely described in terms of a small number of parameters—Fourier coefficients or any other independent descriptors—than wide-band images. Poor human performance with narrow-band images may reflect the impoverished image rather than an intrinsically human characteristic—an ideal observer would exhibit a similar loss.

The problem of assessing the utility of stimulus information becomes acute in comparing human performance in high and in low frequency bandpass filtered images. Typically, filters are constructed to have a bandwidth proportional to frequency (constant bandwidth in terms of octaves). For example, Ginsburg (1980) used faces filtered into 2-octave-wide bands; while Norman and Ehrlich (1987) also used 2-octave bands for their filtered tank pictures. With such filters, high spatial frequency images contain more independent frequencies than low frequency images.

Although linear bandwidth represents perhaps the important difference between images filtered in octave bands at different frequencies, the informational content of the various bands also depends critically on the nature of the specific class of objects, such as faces or letter. Obviously, determining the information content of images is a difficult problem. When it is not solved, the amount of stimulus information available within a frequency band is confounded

with the ability of human observers to use the information. Direct comparisons of performance between differently filtered objects are inappropriate. This distinction between objectively available stimulus information and the human ability to use it has not been adequately posed in the context of spatial bandpass filtering.

Efficiency

In the present context, physically available information is best characterized by the performance of an ideal observer. If there were no noise in the stimulus, the ideal observer would invariably respond perfectly. To compare the performance of an observer, human or ideal, noise of root-mean-square (r.m.s.) amplitude n is progressively added to the signal of r.m.s. amplitude s until the performance is reduced to some criterion, such as 50% correct in a letter identification task. This defines the signal to noise ratio, $(s/n)_c$, for a criterion c. Efficiency *eff* of human performance is defined by:

$$eff = \left(\frac{s_i}{n_i}\right)_c^2 \Big/ \left(\frac{s_h}{n_h}\right)_c^2$$

where h and i indicate *human* and *ideal* observers, and s and n are r.m.s. signal and noise amplitudes (Tanner & Birdsall, 1958). In a pure, quantally limited system, efficiency actually represents the fraction of quanta absorbed (utilization efficiency). In the context of signal detection theory, efficiency is given by a d' ratio:

$$eff = (d'_h/d'_i)^2.$$

Overview

For an object that contains a broad spectrum of spatial frequencies, object spatial frequency is determined by the center frequency of a spatial bandpass filtered image. Retinal spatial frequency is determined by the viewing distance at which the stimulus is viewed. Stimulus information is determined jointly by the signal-to-noise ratio, by the spatial filtering, and by the characteristics of the set of signals; these three informational components are combined in the efficiency computation. Letters are a convenient stimulus to study because they are highly overlearned so that human performance can be expected to be reasonably efficient, and because much is already known about the visibility of letters in the presence of internal noise (letter acuity) and about the visual processing of letters.

Specifically, to determine the roles of object and retinal spatial frequencies, letters are filtered into various frequency bands. Noise is added, and the psychometric function for correct identification is determined as a function of s/n. Accuracy depends only on s/n and not on overall contrast, for a wide range of contrasts (Pavel, Sperling, Riedl & Vanderbeck, 1987). This determination is repeated for every combination of object frequency band and viewing distance. Thereby, retinal spatial frequency and object spatial frequency are unconfounded, enabling us to determine whether a particular object frequency band is better discriminated in one visual channel (retinal frequency) than any other (Parish & Sperling, 1987a, b). Moreover, by computing an ideal observer for the identification task, we obtain an objective measure of the information that is present in each of the frequency bands. Finally, the comparison of human performance with the performance of the ideal observer gives us a precise measure of the ability of our subjects to utilize the information in the stimulus. Having untangled these factors, we can determine which spatial frequencies most efficiently characterize letters for identification.

METHOD

Two experiments were conducted using similar stimuli and procedures.

Stimuli

Letters (signals) and noise. The original, unfiltered letters were selected from a simple 5 × 7 upper-case font commonly used on CRT terminals. Since this is an experiment in pattern recognition, we felt that the simplest letter pattern might be the most general; indeed, this font has been widely used in letter discrimination studies. For the purpose of subsequent spatial filtering, the letters were redefined on a pixel grid that measured 45 (vertical height) × 35 (maximum horizontal extent of letters M and W). The letters had value 1 (white); the background had value 0 (black). To avoid edge effects in filtering, the background was extended to 128 × 128 pixels for all computations. However, only the center 90 × 90 pixels of the stimulus were displayed, as these contained effectively all the usable stimulus information, even for low spatial-frequency stimuli. Letters for presentation were chosen pseudo-randomly from the set of 26 upper-case English letters. Noise

Table 1. Parameters of the bandpass filters: lower and upper half-amplitude frequencies, peak, and 2D mean frequencies in cycles/letter height

Band	Lower	Peak	Upper	Mean[a]
0	0	Lowpass	0.53	0.39
1	0.26	0.53	1.05	0.74
2	0.53	1.05	2.11	1.49
3	1.05	2.11	4.22	2.92
4	2.11	4.22	8.44	5.77
5	6.33	Highpass	22.5	20.25

[a]Frequencies are weighted according to their squared amplitude (power) in computing the mean.

fields were defined on a 128 × 128 array by choosing independent Gaussian noise samples for each pixel, with the mean equal to zero and a variance σ^2 as required by the condition. (As with the letters, only the central 90 × 90 pixels were displayed.) Forty different noise fields were created.

Filters. Each stimulus consisted of a filtered letter added to an identically filtered noise field. Six spatial filters were available, corresponding to six successive levels of a Laplacian pyramid (Burt & Adelson, 1983). The zero-frequency component was added to the images so that they could be viewed. The object-relative filter characteristics, upper and lower half-amplitude cutoff and 2D mean frequency (cycles per letter height), appear in Table 1. The 2D mean frequency $\bar{f}$ for a given band is:

$$\bar{f} = \sum_{x=0}^{127}\sum_{y=0}^{127} f_{x,y}a_{x,y}^2 \Big/ \sum_{x=0}^{127}\sum_{y=0}^{127} a_{x,y}^2$$

where $f_{x,y}$ is the 2D frequency and $a_{x,y}$ is its amplitude. Cycles per object height is used rather than the more usual cycles per object width because the height of our upper-case letters remained constant across the entire set, whereas the width varied between letters.

The transfer functions (spectra) of the filters are displayed in Fig. 1. Approximately, filters are separated in spatial frequency by an octave (factor of 2) and have a bandwidth at half-amplitude of two octaves. The small mound in the lower right corner of Fig. 1 is a negligible imperfection in filter 4. For convenience, the limited range of spatial frequencies passed by each of the filters will be referred to as the *band* of that filter; a specific band is b_i ($i = 0, 1, 2, 3, 4, 5$), where b_0 is the lowest set of frequencies and b_5 is the highest.

The filter spectra (shown in Fig. 1) are approximately symmetrical in log frequency coordinates, a symmetrical spectrum in log coordinates is highly skewed to the right in linear frequency coordinates, resulting in a mean that

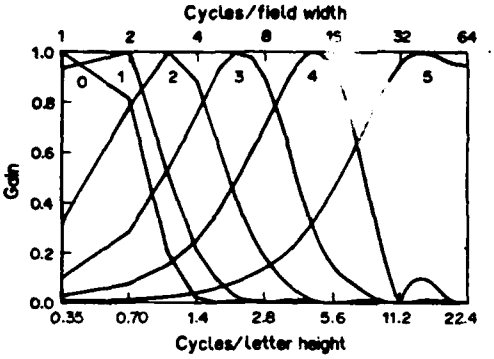


Fig. 1. Filter characteristics for the filters used in the experiments. There are two abscissas, both on a log scale. The top abscissa is the frequency in cycles per unwindowed field width (128 pixels); the bottom abscissa is in cycles per letter height (45 pixels). The ordinate is the normalized gain. The parameter i indicates the filter designation b_i in the text.

is much greater than the mode. In a 2D (vs 1D) filter, the rightward shift is accentuated. For example, band 2 has a peak frequency of 1.05 c/object but a 2D mean frequency of 1.49 c/object. The single most informative characterization of such a skewed bandpass spectrum depends somewhat on the context; usually use the mean rather than the peak.

Figure 2 (top) shows the letter G, filtered in bands 1–5 without noise; the bottom shows the same signals plus noise, $s/n = 0.5$. The full 128 × 128 array (extended by reflection beyond its edges) was passed through the filter so that the effect of the picture boundary did not intrude into the critical part of the display.

Signal to noise ratio, s/n. A filtered letter is a *signal*. Let i, j index a particular pixel in the x, y coordinate space of the stimulus. The signal contrast $c_s(i,j)$ of pixel i,j is:

$$c_s(i,j) = \frac{(l(i,j) - l_0)}{l_0} \qquad (1)$$

where $l_{i,j}$ is the luminance of pixel i, j and l_0 is the mean signal luminance over the 90 × 90 array. Signal power per pixel, s, is defined as mean contrast power averaged over the 90 × 90 pixel array:

$$s = (IJ)^{-1}\sum_{i}^{I}\sum_{j}^{J} c_s(i,j)^2 \qquad (2)$$

where $c_{i,j}$ is the contrast of pixel i, j and $I = J = 90$.

Noise contrast $c_n(i,j)$ is the value of the i,jth noise sample divided by the mean luminance. Analogously to signal power (equation 2), noise contrast power per pixel, n, is equal to $(\sigma/l_0)^2$. The signal to noise ratio is simply s/n.

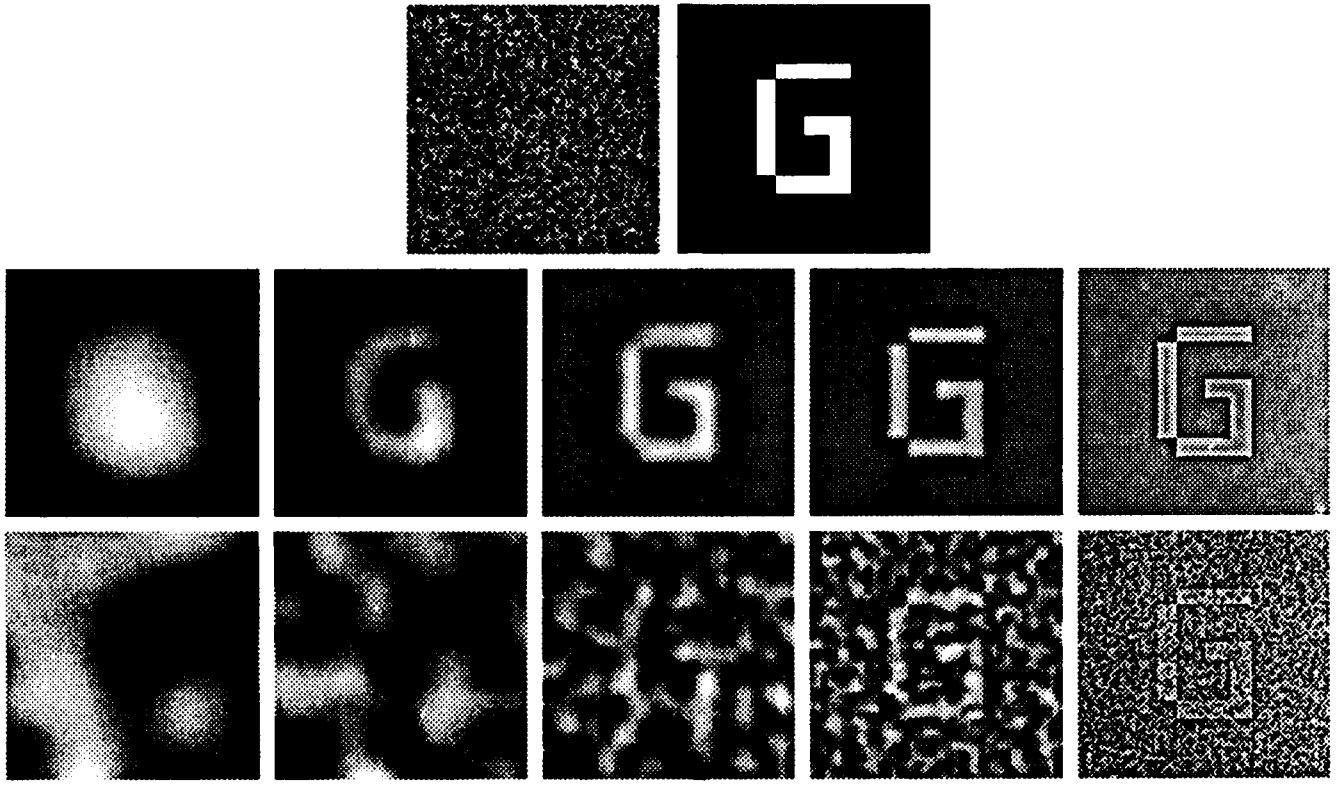


Fig. 2. Top: unfiltered noise and unfiltered letter G. Middle: the letter G filtered in spatial frequency bands 1–5 with only quantization noise. Bottom: filtered letter G plus filter noise in the same bands with a signal-to-noise ratio of 0.50 in all panels. The effective s/n in the reproduction is somewhat lower (from Parish & Sperling, 1987a). The first row of numerals indicates the number by which the filter band is referred to in the text; the bottom row indicates the *mean* frequency of the bands in cycles per letter height.

Quantization. Our display system produced 256 discrete luminance levels. Level 128 was used as the mean luminance l_0; l_0 was 47.5 cd/m^2. To produce a visual display of a given letter, band, and s/n, signal power s and noise power n were normalized so that the luminance of every one of the 8100 displayed pixels fell within the range of the display system; there was no truncation of the tails of the Gaussian noise. (Although the relationship between input gray-level and output luminance was not quite linear at the extreme intensity values, it was determined that more than 90% of the pixels fell within the linear intensity range.) Intensity normalization was applied separately to each stimulus (combination of signal plus noise). By normalizing the total stimulus $s + n$, the actual value of s displayed to the subject diminished as n increased; i.e. the actual value of s was not known by the subject. Indeed, even stimuli with precisely the same letter in the same band and with the same s/n might be produced with slightly different s and n depending on the extreme values of the noise fields.

Seven values of s/n were available for each band, chosen in a pilot study to insure that the data yielded the entire psychometric function (chance to best performance). The same pilot study showed that subjects never performed above chance when confronted with noise-free letters from b_0; this band was omitted from the present study.

Procedure: experiment 1

Four of the experimental variables—letter identity, noise field, frequency band, and s/n—were randomized within each session. A fifth variable, viewing distance, was held constant within each session and was varied between sessions. Four viewing distances were used: 0.121, 0.38, 1.21 and 3.84 m. A chin rest was used to stabilize the subject's head for viewing at the shortest distance. At the four distances, the 90 × 90 pixel stimulus subtended 31.6, 10, 3.16 and 1.0 deg of visual angle respectively. The upper and lower half-amplitude cut-off retinal frequencies for the upper six filters, with respect to the four viewing distances used in this experiment, and for a fifth distance used in the second experiment, appear in Table 2. Subjects participated in four 1-hr sessions at each viewing distance. Each session consisted of 315 trials, nine trials at each of seven s/n's for each of the five frequency bands.

Prior to the first session, subjects were shown noise-free examples of the unfiltered letters. They were told that each stimulus presentation consisted of a letter and a certain amount of noise, and that the letter may appear degraded in some way. They were informed that at no time would a letter be shifted in orientation or from its central location in the stimulus field. Finally, they were instructed to view each stimulus for as long as they desired before making their best guess as to which letter had been presented. A response (letter identity) was required on every trial. Subjects typed the response on a keyboard connected to the host computer (Vax 11/750); subsequently, typing a carriage return erased the video screen and initiated the next trial in a few seconds. The room illumination was very dim; the response keyboard was lighted by stray light from its associated CRT terminal. No feedback was offered to the subjects.

Observers

Three subjects, two male and one female, between the ages of 20 and 27 participated in the experiment. All subjects had normal or corrected-to-normal vision. One of the subjects was a paid participant in the study.

Procedure: experiment 2

This experiment was run before expt 1. It is reported here because it offers additional data with two new and one old subject at a fifth viewing distance. Except as noted, the procedures are similar to expt 1. The screen was viewed through a darkened hood at a distance

Table 2 Lower and upper half-power frequency and 2D mean frequency (in c/deg of visual angle) for all bands and viewing distances used in both experiments

Band	Viewing distance (m) 0.12	0.38	1.21	3.84	0.48
0 (lowpass)	0.00–0.04 (0.03)	0.00–0.12 (0.09)	0.00–0.37 (0.27)	0.00–1.18 (0.87)	0.00–0.15 (0.11)
1	0.02–0.07 (0.05)	0.06–0.23 (0.16)	0.18–0.74 (0.52)	0.58–2.34 (1.65)	0.07–0.29 (0.21)
2	0.04–0.15 (0.10)	0.12–0.47 (0.33)	0.37–1.48 (1.04)	1.18–4.70 (3.30)	0.15–0.59 (0.41)
3	0.07–0.30 (0.20)	0.23–0.94 (0.64)	0.74–2.97 (2.04)	2.34–9.40 (6.48)	0.29–1.18 (0.81)
4	0.15–0.59 (0.40)	0.47–1.88 (1.27)	1.48–5.94 (4.04)	4.70–18.80 (12.82)	0.59–2.36 (1.60)
5 (highpass)	0.30–2.25 (1.41)	0.94–7.13 (4.45)	2.97–22.53 (14.19)	9.40–71.27 (45.00)	1.77–8.96 (5.63)

of 0.48 m. At this distance, the 90 × 90 stimuli subtended 7.15 deg of visual angle. The half-amplitude cut-off frequencies and the mean frequencies of the six spatial filters are given in the rightmost column of Table 2. Three male subjects between the ages of 20 and 27 participated in the experiment. All subjects had normal or corrected-to-normal vision. Two of the subjects were paid for their participation, and one, DHP, also participated in expt 1. Five sessions of 315 trials were run for each subject.

RESULTS

Psychometric functions: $\hat{p}$ vs $\log_{10} s/n$

The measure of performance is the observed probability $\hat{p}$ of a correct letter identification. The complete psychometric functions are displayed in Figs 3 (expt 1) and 4 (expt 2). A separate psychometric function is shown for each subject, viewing distance and frequency band. In band b_1, for all subjects, performance asymptotes (for noiseless stimuli) at $\hat{p} \approx 0.5$. In all other bands, performance improves from near-chance (1/26) to near perfect as the value of s/n increases.

Noise resistance as a function of frequency band

An obvious aspect of the data of both experiments is that the data move to the left of the figure panels as band spatial frequency increases. This means that high spatial frequency stimuli (bands b_4, b_5) are identifiable at smaller

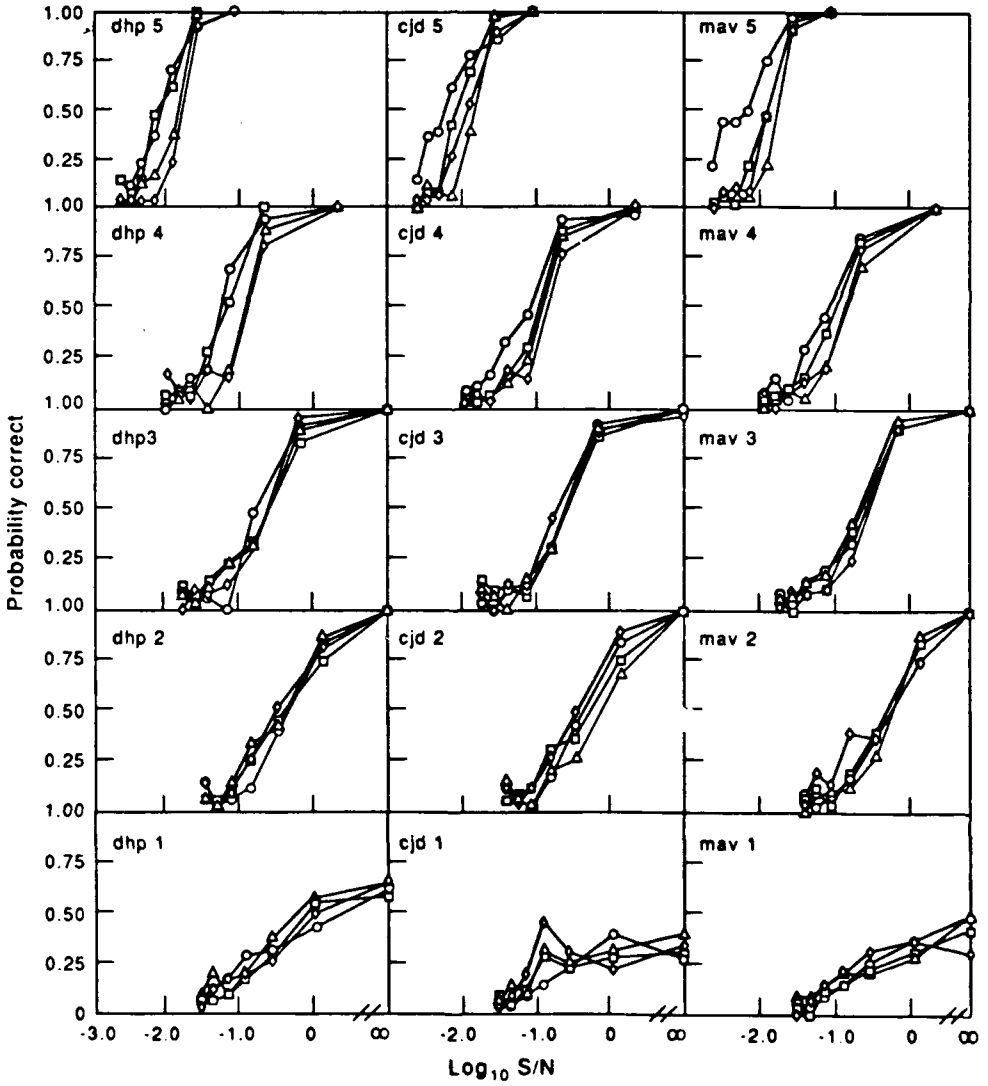


Fig. 3. Psychometric functions from expt 1. Each graph displays performance as a function of $\log_{10} s/n$, within a frequency band. The parameter is viewing distance. Subjects are arranged in columns and frequency band is arranged in rows, progressing from the highest frequency band at the top to the lowest band at the bottom. The four viewing distances are 3.84 (○), 1.21 (△), 0.38 (□), and 0.121 (◇ m

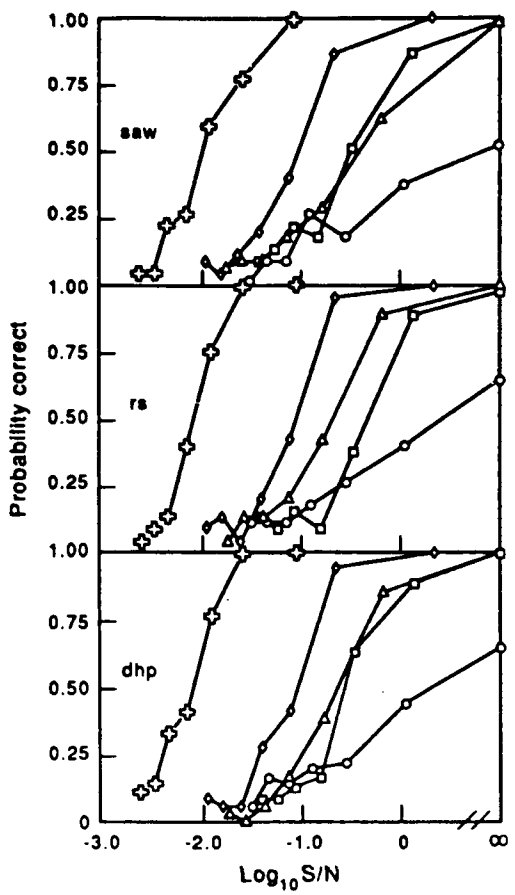


Fig. 4. Psychometric functions for each subject and frequency band in expt 2. Viewing distance was 0.48 m. The five frequency bands, b_1–b_5, are indicated, respectively, by ○, □, △, ◇ and +. The probability of a correct response is plotted as a function of $\log_{10} s/n$.

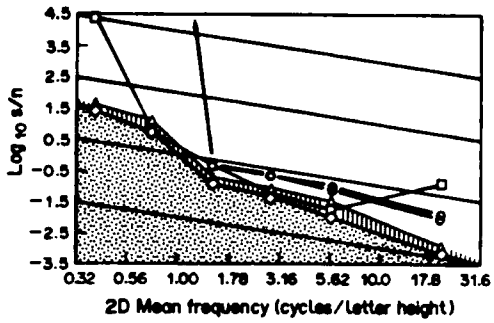


Fig. 5. Performance of human subjects and various computational discriminators. The abscissa indicates $\log_{10}$ of the mean frequency of each bandpass stimulus. The ordinate indicates the (interpolated) s/n ratio at which a probability of a correct response $p = 0.5$ is achieved. Circles indicate each of the three subjects in expt 1 at the intermediate viewing distance of 1.21 m. In band b_1, 2 of 3 human subjects fail to achieve 50% correct ($eff = 0$); these points lie outside the graph. (△) indicates sub-ideal and (◇) indicates super-ideal performances of discriminators that brackets the ideal discriminator. The shaded area below the super-ideal discriminator indicates theoretically unachievable performance. Squares indicate performance of a spatial correlator-discriminator. The oblique parallel lines have slope −1 that represents the improvement in expected performance (decrease in s/n) as function of the number of frequency components in each band when filter bandwidth is proportional to frequency.

s/n than stimuli in bands b_1 and b_2; resistance to noise increases with spatial frequency band. To enable comparisons of noise sensitivity as a function of band, the s/n at which $\hat{p} = 50\%$ was estimated for each subject and frequency band from expt 1 by means of inverse interpolation from the best fitting logistic function. As viewing distance had no effect, all estimates were made using the data collected when viewing distance was equal to 0.38 m. A graph of these $(s/n)_{50\%}$ points as a function of the mean object frequency of the band is plotted in Fig. 5 (○). For comparison, the expected rate of improvement in $(s/n)_{50\%}$, based on the increasing number of frequency components as one moves from low to high frequency bands, is plotted as a series of parallel lines in Fig. 5. Performance improves [$(s/n)_{50\%}$ decreases] somewhat faster than $1/f$ (the slope of the parallel lines). These results, and Fig. 5, will be analyzed in detail in the Discussion section.

The non-effect of viewing distance

Another property of the data is that, in most conditions, viewing distance has no effect on performance. Analysis of variance, carried out individually for each subject, shows that there is no significant effect of distance in any band for subject dhp and a significant effect of distance in bands b_4 and b_5 for the other two subjects. Further analysis by a Tukey test (Winer, 1971) in bands b_4 and b_5 for these subjects shows that the only significant effect of distance is that visibility at the longest viewing distance is *better* than at the other three distances. For subject CJD, the improvement is equivalent to a gain in s/n of 0.19 and 0.28 $\log_{10}$ (for bands b_4 and b_5, respectively); for MAV, the corresponding gains were 0.21 and 0.40.

Improved performance at long viewing distances is almost certainly due to the square configuration of individual pixels, which produces a high frequency spatial pixel noise that is attenuated by viewing from sufficiently far away (Harmon & Julesz, 1973). In low frequency bands, pixel-boundary noise is not a problem because the spatial filtering insures that adjacent pixels vary only slightly in intensity. We explored the hypothesis of pixel-boundary noise with subject CJD, who showed a distance effect

in band 5. At an intermediate viewing distance of 1.21 m, CJD squinted her eyes while viewing stimuli from band 5. By blurring the retinal image of the display in this way, performance improved approximately to the level of the furthest viewing distance.

To summarize, the only significant effect of distance that we observed was a lowering of performance at near viewing distances relative to the furthest distance. This impairment occurred primarily in bands 4 and 5. In these bands, the spatial quantization of the display (90 × 90 square-shaped pixels) produces artifactual high spatial frequencies that mask the target. These artifactually produced spatial frequencies can be attenuated by deliberate blurring (squinting), or by producing displays with higher spatial resolution, or by increasing the viewing distance to the point where the pixel boundaries are attenuated by the optics of the eye and neural components of the visual modulation transfer function. In all cases, blurring improves performance and eliminates the slightly deleterious effect of a too small viewing distance. Thus, for correctly constructed stimuli, in the frequency ranges studied, there would be no significant effect of viewing distance on performance. This finding is in agreement with the results of Legge et al. (1985), who examined reading rate rather than letter recognition. It is in stark disagreement with the results of sinewave detection experiments in which retinal frequency is critical—see Sperling (1989) for an explanation.

DISCUSSION

A comparison of performance in different frequency bands shows that subjects perform better the higher the frequency band; and subjects require the smallest signal-to-noise ratio in the highest frequency band. To determine whether performance in high frequency bands is good because humans are more efficient in utilizing high-frequency information, or because there is objectively more information in the high-frequency images, or both, requires an investigation of the performance of an ideal observer. The performance of the ideal observer is the measure of the objective presence of information. Human performance results from the joint effect of the objective presence of information and the ability of humans to utilize that information. Human efficiency is the ratio of human performance to ideal performance.

Ideal discriminator

Definition. An ideal discriminator makes the best possible decision given the available data and the interpretation of "best." The performance of the ideal discriminator defines the objective utility of the information in the stimulus. We prefer the name *ideal discriminator*, rather than *ideal observer*, because it indicates the critical aspect of performance under consideration, but we occasionally use *ideal observer* to emphasize the relations to a large, relevant literature on this subject. Our purposes in this section are first, to derive an ideal discriminator for the letter identification task, second, to develop a practical working approximation to this discriminator, and third, to compare the performance of the human with the ideal discriminator.

Although ideal observers have recently come into greater use in vision research, the applications have focused primarily on determining the limits of performance for relatively low-level visual phenomena. For example, Barlow (1978, 1980), and Barlow and Reeves (1979) investigated the perception of density and of mirror symmetry; Geisler (1984) investigated the limits of acuity and hyperacuity; Legge, Kersten and Burgess (1987) examined the pedestal effect; Kersten (1984) studied the detection of noise patterns; and Pelli (1981) detailed the roles of internal visual noise. Geisler (1989) provides an overview of efficiency computations in early vision. Our application differs from these in that we expand the techniques and apply them to a higher perceptual/cognitive function, letter recognition.

For the letter identification task, the ideal discriminator is conceptually easy to define. A particular observed stimulus, x, representing an unknown letter plus noise, consists of an intensity value (one of 256 possible values) at each of 90 × 90 locations. The discriminator's task is to make the correct choice as frequently as possible from among the 26 alternative letters.

The likelihood of observing stimulus x, given each of the 26 possible signal alternatives, can be computed when the probability density function of the added noise is known exactly. The optimal decision chooses the letter that has the highest likelihood of yielding x. The expected performance of the ideal discriminator is computed by summing its probability of a correct response over the 256^{8100} possible stimuli (256 gray levels, 90 × 90 pixels). Unfortunately,

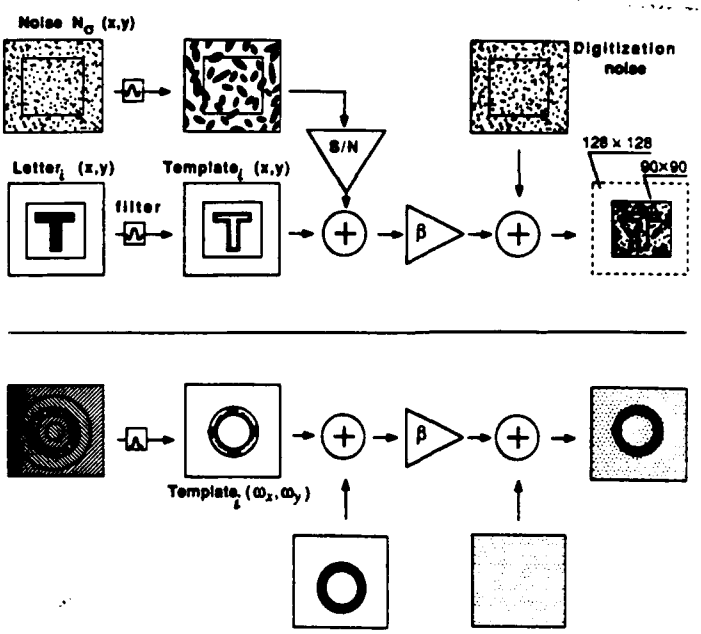


Fig. 6. Flow chart of the experimental procedures that are modelled by the ideal discriminator analysis. Upper half indicates space-domain operations; lower half indicates the corresponding operations in the frequency domain. Computations are carried out on 128 × 128 arrays; the subject sees only the center 90 × 90 pixels. A random letter and a random noise field are each filtered by the same filter (b); the noise is amplified to provide the desired signal-to-noise ratio; the letter and noise are added, the output is scaled and quantized (represented by the addition of digitization noise), and the result is shown to the subject. In the frequency domain ω_x, ω_y, the bandpass filter selects an annulus, whereas the quantization noise is uniform over ω_x, ω_y.

when there is both bandpass filtered and intensity quantization, the usual simplifications that make this enormous computation tractable are not applicable.

As an alternative to computing the expected performance of the ideal discriminator, one can compute its performance with a particular subset of the possible stimuli—the stimuli that the subject actually viewed or, preferably, a larger set of stimuli for more reliable estimation. This Monte Carlo simulation of the performance of the ideal discriminator is a tractable computation that yields an estimate of expected performance.

Derivation. Stimulus construction is diagrammed in Fig. 6 which shows the equivalent operations in the space and the frequency domains. To derive an ideal discriminator, we need to carefully review the processes of stimulus construction. We use uppercase letters to represent quantities in the frequency domain and lowercase letters to represent quantities in the space domain. A letter is defined by a 90 × 90 array that takes the value 1 at the letter locations and 0 at the background locations. When this array is spatially filtered in band b, it defines the *letter template* $t_{i,b}(x, y)$, where i indicates the particular letter, b the frequency band, and x, y the pixel location. We write $T_{i,b}(\omega_x, \omega_y)$ for the Fourier series coefficient of $t_{i,b}$ indexed by frequency.

An unknown stimulus $u_{i,b}(x, y)$ to be viewed by a subject is produced by adding filtered $n_b(x, y)$ with post-filtering variance σ_N^2, to the template $t_{i,b}(x, y)$, where letter identity i is unknown to the subject. The stimulus is scaled and digitized (quantized) to 256 levels prior to presentation, contributing an additional source of noise $q_{i,b}(x, y)$, called digitization noise. Finally, a d.c. component (dc) is added to $u_{i,b}$ to bring the mean luminance level to 128. These steps are diagrammed in Fig. 6 which shows both the space-domain and the corresponding frequency-domain operations. The space-domain computation is encapsulated in equations (3):

$$u_{i,b}(x, y) = \beta_{i,b}[t_{i,b}(x, y) + n_b(x, y)] \quad (3a)$$

$$u_{i,b}(x, y) = \beta_{i,b}[t_{i,b}(x, y) + n_b(x, y)] + q_{i,b}(x, y) + dc. \quad (3b)$$

The scaling constant $\beta_{i,b}$, limits the range of real values for each pixel, prior to quantization, to $[-0.5, 255.5]$. The degree of scaling is determined by the maximum and minimum values in

the function $t_{i,b} + n_b$. Note that the extreme values in the image are determined by σ_{N_2}, which is adjusted to yield the appropriate s/n for each condition; the values of $t_{i,b}$ are fixed prior to scaling. Specifically:

$$\beta_{i,b} = \frac{256}{\max(t_{i,b} + n_b) - \min(t_{i,b} + n_b)}. \quad (4)$$

As a result of bandpass filtering, the noise samples in adjacent pixels are strongly dependent on each other. Therefore, the discriminator problem is best approached in the Fourier domain, where the random variables $\{N_b(\omega_x, \omega_y)\}$ are jointly independent because the filtering operations simply scale the different frequency components without introducing any correlations (van Tress, 1968). The task of the ideal discriminator is to pick the template $t_{i,b}$ that maximizes the likelihood of $u_{i,b}$ with *a priori* knowledge of: (i) the fixed functions $t_{i,b}$, and their probabilities; and (ii) the densities of the jointly independent random variables $\{N_b(\omega_x, \omega_y)\}$. As is clear, $\beta_{i,b}$, σ_N^2, $\{Q_{i,b}(\omega_x, \omega_y)\}$, and $\{N_{i,b}(\omega_x, \omega_y)\}$ are all jointly distributed random variables characterized by some density f. To compute the likelihood of $u_{i,b}$ the ideal discriminator must integrate f over all possible values that may be assumed by the set of jointly distributed random variables, whose values are constrained only in that they result in a possible stimulus $u_{i,b}$. Unfortunately, no closed-form solution to this problem is available, forcing us to look for an alternative approach.

Bracketing. To estimate the performance of the ideal discriminator, we look for a tractable super-ideal discriminator that is better than the ideal but which is solvable. Similarly, we look for a tractable sub-ideal discriminator that is worse than the ideal. The ideal discriminator must lie between these two discriminators; that is, we bracket its performance between that of a "super-ideal" and a "sub-ideal" discriminator. The more similar the performance of the super- and sub-ideal discriminators, the more constrained is the ideal performance which lies between them.

Our super-ideal discriminator is told, *a priori*, the exact values for $\beta_{i,b}$ and σ_N^2 for each stimulus presentation. Therefore, it is expected to perform slightly better than the ideal discriminator which must estimate these values from the data. The sub-ideal discriminator estimates these same parameters from the presented stimulus in a simple but nonideal way. Therefore, it is expected to perform slightly worse than the ideal discriminator. The computational forms used to compute $\beta_{i,b}$ and σ_N^2 for the sub-ideal discriminator are presented in the Appendix, along with the derivation of the likelihood estimator used by both discriminators. A complete discussion of these derivations and the problems associated with the formulation of an ideal discriminator for such complex stimuli is presented in Chubb, Sperling and Parish (1987).

Performance of the bracketed discriminator. The super- and sub-ideal discriminators were tested in a Monte Carlo series of trials, in which they each were confronted with 90 stimuli in each of the frequency bands at each of seven s/n values chosen to best estimate their 50% performance point. The s/n necessary for 50% correct discriminations was estimated by an inverse interpolation of the best fitting logistic function. The derived $(s/n)_{50\%}$ is the measure of performance of a discriminator. The mean ratio, across frequency bands, of

$$(s/n)_{50\%} \text{ sub-ideal}/(s/n)_{50\%} \text{ super-ideal}$$

is about 2 (approx. 0.3 $\log_{10}$ units). The ratio does not depend on the criterion of performance.

Efficiency of human discrimination

In all conditions, human subjects perform worse than the sub-ideal discriminator. Notably, with no added luminance noise, the subideal (and, of course, the ideal) discriminator function perfectly, even in b_0 where subject performance is at chance, and in b_1 where subject reached asymptote at about 50% correct.

Data from the subjects are plotted with the $(s/n)_{50\%}$ sub-ideal and $(s/n)_{50\%}$ super-ideal in Fig. 5. For comparison, Fig. 5 also shows the performance of a correlator discriminator which chooses the letter template that correlates most highly with the stimulus in the space domain. In the coordinates of Fig. 5 ($\log_{10} s/n$ vs $\log_{10} f$ where f represents the mean 2D spatial frequency of the band), the vertical distance d from the human data $\log(s/n)_{50\%}$, *human* down to the bracketed discriminator $\log(s/n)_{50\%}$, *ideal* represents the $\log_{10}$ of the factor by which the bracketed discriminator outperforms the human observer at that value of f. For the purpose of specifying efficiency, we assume the ideal discriminator lies at the mid-point of the sub and super-ideal discriminators in Fig. 5. The

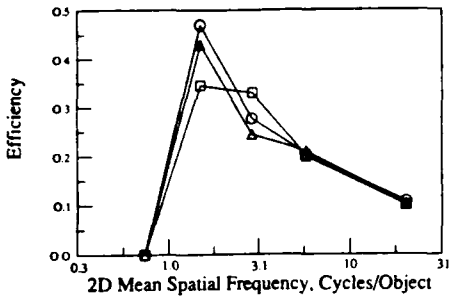


Fig. 7. Discrimination efficiency as a function of the mean frequency of a 2-octave band (in cycles per letter height) indicated on a logarithmic scale. Data are shown for three observers: △ = SAW, □ = RS, ○ = DHP. The viewing distance is 2.21 m, which is representative of all viewing distances tested.

efficiency *eff* of human discrimination relative to the bracketed discriminator is $eff = 10^{-2d}$, where:

$$d = \log(s/n)_{50\%,\,human} - \log(s/n)_{50\%,\,ideal}.$$

The values of *eff* in each object frequency band are shown in Fig. 7. In band 0, *eff* is zero because human performance never reaches 50%; indeed, it never rises significantly above 4% (chance). In band 1, human performance asymptotically climbs close to 50% as s/n approaches infinity; $eff \approx 0$. In band 2, *eff* reaches its maximum of 35–47% (depending on the subject), and it declines rapidly with increasing frequency (b_3–b_5).

The 42% average efficiency in band 2 is similar in magnitude to the highest efficiencies observed in comparable studies. For example, efficiency has been determined for detecting various kinds of patterns in arrays of random dots (Barlow, 1978, 1980; van Meeteren & Barlow, 1981), tasks which, like ours, may require significantly cognitive processing. In a wide range of conditions, the highest efficiencies observed were about 50%, and frequently lower. Van Meeteren and Barlow (1981) also found that efficiency was perfectly correlated with object spatial frequency and was independent of retinal spatial frequency.

Spatial correlator discriminator. A correlator discriminator cross-correlates the presented stimulus with its memory templates and chooses the template with the highest correlation. Correlation can be carried out in the space or in the frequency domain. Correlation is an efficient strategy when noise in adjacent pixels is independent and when members of the set of signals have the same energy; both of these conditions are violated by our stimuli. However, when sufficient prior information is available to subjects, they do appear to employ a cross-correlation strategy (Burgess, 1985).

It is interesting to note that the performance of the spatial correlator discriminator over the middle range of spatial frequencies is quite close to the performance of the sub-ideal discriminator. At high spatial frequencies, correlator performance degenerates, due to its inability to focus spatially on those pixel locations that contain the most information. A spatial correlator that optimally weighted spatial locations, could overcome the spatial focusing problem at high frequencies. (Spatial focusing is treated in the next section.)

At all frequencies, the spatial correlator is nonideal because noise at spatial adjacent pixels is not independent. At low spatial frequencies, the nonindependence of adjacent locations becomes extreme and the correlator fails miserably. This points out that, for our stimuli, correlation detection is better carried out in the frequency domain because there the noise at different frequencies is independent. The qualitative similarity between the correlator discriminator and the subjects' data suggests that the subjects might be employing a spatial correlation strategy, augmented by location weighting at high frequencies.

Lowest spatial frequencies sufficient for letter discrimination. Band 2 corresponds to a 2-octave band with a peak frequency of 1.05 c/object (vertical height of letters) and a 2D mean frequency of 1.49 c/object. At the four viewing distances, 1.05 c/object corresponds to retinal frequencies of 0.074, 0.234, 0.739 and 2.34 c/deg of visual angle. We observe perfect scale invariance: all of these retinal frequencies, and hence the visual channels that process this information, are equally effective in achieving the high efficiency of discrimination.

The finding that b_2 with a center frequency of 1.05 c/object and a $\frac{1}{2}$ amplitude cutoff at 2.1 c/object is critical for letter discrimination is in good agreement with previous findings of both Ginsburg (1978) for letter recognition and Legge et al. (1985) for reading rate. Legge et al. used low-pass filtered stimuli, which included not only spatial frequencies within an octave of 1 c/object (b_2) but also included all lower frequencies. From the present study, we expect human performance with low-pass and with band-pass spatial filtering to be quite similar up to 1 c/object because the lowest frequency

bands, when presented in isolation, are perceptually useless (at least when presented alone).

It is an important fact that our subjects actually performed better, in the sense of achieving criterion performance at a lower s/n ratio, at higher frequency bands than b_2. This is explained by the increase in stimulus information in higher frequency stimuli. Increased information more than compensates for the subjects' loss in efficiency as spatial frequency increases.

Components of discrimination performance

Though the performance of the bracketed ideal discriminator is useful in quantifying the informational utility of the various bands, it is instructive to consider the changing physical structure of the stimuli as well. What components of the stimuli actually lead to a gain in information with increasing frequency? According to Shannon's theorem (Shannon & Weaver, 1949), an absolutely bandlimited 1-D signal can be represented by a number of samples m that is proportional to its bandwidth. When the signal-to-noise ratio in each sample s_i/n_i is the same, the overall signal-to-noise ratio s/n grows as $\sqrt{m}$. In the space domain, our filters were constructed (approximately) to differ only in scale but not in the shape of their impulse responses. Therefore, when the mean frequency of a filter band increased by a factor of 2, the bandwidth also increased by 2. Since the stimuli are 2D, the effective number of samples increases with the square of frequency, and the increase in effective s/n ratio is proportional to m. This expected improvement with frequency, based simply on the increase in effective number of samples, is indicated by the oblique parallel lines of Fig. 5 with slope of -1. The expected improvement in threshold s/n due simply to the linearly increasing bandwidth of the bands does a reasonable job of accounting for the improvement in performance for both human and bracketed discriminators between b_2 and b_5.

Performance of all discriminators improves faster with frequency between 0.39 and 1.5 c/object and between 5.8 and 22 c/object than is predicted from the bandwidths of the images. A slope steeper than -1 means that there is more information for discriminating letters in higher frequency bands even when the number of independent samples is kept the same in each band. Once sampling density is controlled, just how much information letters happen to contain in each frequency band is an ecological property of upper-case letters.

Increasing spatial localization with increasing frequency band. From the human observer's point of view, the letter information in low-pass filtered images is spread out over a large portion of the total image array. In high spatial-frequency images, the letter information is concentrated in a small proportion of the total number of pixels. In high spatial-frequency images, a human observer who knows which pixels to attend will experience an effective s/n that is higher than an observer who attends equally to all pixels. In this respect, humans differ from an ideal discriminator. The ideal discriminator has unlimited memory and processing resources, does not explicitly incorporate any selective mechanism into its decision, and uses the same algorithm in all frequency bands. Information from irrelevant pixels is enmeshed in the computation but cancels out perfectly in the letter-decision process. To understand human performance, however, it is useful to examine how, with our size-scaled spatial filters, letter information comes to be occupy a smaller and smaller fraction of the image array as spatial frequency increases.

Here we consider three formulations of the change in the internal structure of the images with increasing spatial frequency: (1) spatial localization; (2) correlation between signals; and (3) nearest neighbor analysis. We have already noted that, in our images, the information-rich pixels become a smaller fraction of the total pixels as frequency band increases. Indeed, this reduction can be estimated by computing the information transmitted at any particular pixel location or, more appropriately for estimating noise resistance, by computing the variance of intensity (at that pixel location) over the set of 26 alternative signals.

To demonstrate the degree of increasing localization with increasing frequency, the variance (over the set of 26 letter templates) was computed at each pixel location (x, y). *Total power*, the total variance, is obtained by summing over pixel locations. The number of pixel locations needed to achieve a specific fraction of the total power is given in Fig. 8, with frequency band as a parameter. These curves describe the spatial distribution of information in the latter templates. If all pixels were equally informative, exactly half of the total number of pixels would be needed to account for 50% of the total power. The solid curves in Fig. 8 show that the number of pixels needed to convey any percentage of total signal power, decreases as the

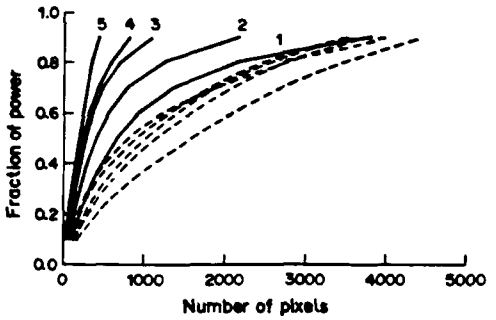


Fig. 8. Fraction of total power contained in the n most extreme-valued pixels as a function of n (out of 8100). Solid lines indicate the power fractions for signals; the curve parameter indicates the filter band. Dashed lines indicate power fractions for filtered noise fields. Although power fractions from successive bands of noise are too close to label, they generally fall in the same left–right 5–0 order as those for signal bands.

frequency band increases. These information distribution curves are an ecological property of our set of letter stimuli; different curves would be needed describe other stimulus sets.

The dashed curves in Fig. 8 were derived from random noise filtered in each of the six frequency bands (b_0–b_5). The distribution of noise power is very similar between the various bands, enormously more so than the distribution of signal power. For our letter stimuli, stimulus information coalesces to a smaller number of spatial locations as spatial frequency increases.

Correlation between signals. A more abstract way of describing the change of information with bandwidth is to note that letters become less confusible with each other in the higher frequency bands. A good measure of confusibility is the average pairwise correlation between the 26 letter templates in each frequency band (Table 3). The average correlation between letter templates diminishes from 0.94 in band 0 to 0.31 in band 5. In a band in which templates have a pairwise correlation over 0.9, the overwhelming amount of intensity variation ("information") is useless for discrimination. Small wonder that subjects fail completely in this band. Overall, performance of the ideal discriminator and of observers improves as the correlation decreases, but there is no obvious way to use the pairwise correlation between templates to predict performance.

Nearest neighbors. The analysis of nearest neighbors is a useful technique for predicting accuracy by the analysis of the possible causes of errors. We can regard a filtered image t_i of letter i as a vector in a space of dimensionality 8100 (90 × 90 pixels). When noise is added, the possible positions of t_i are described by a cloud whose dimensions are determined by the s/n ratio. A neighboring letter k may be confused with letter i when the cloud around t_i envelopes t_k. The closer the neighbor, the greater the opportunity for error. Table 3 gives the average normalized distance to the nearest neighbor in each of the bands. The increase in distance to the nearest neighbor reflects the improvement in the representation of signals as spatial frequency increases.

Table 3. Average pairwise correlations and nearest neighbors (Euclidean distance × 10^{-5})

Band	Correlations	Nearest neighbor
0	0.94	0.01
1	0.91	0.30
2	0.58	1.2
3	0.38	2.3
4	0.33	3.1
5	0.31	4.1

We consider possible causes of lower efficiency of discrimination in bands below b_2. The letters in these bands have high pair-wise correlations and the mean band frequency is less than the object frequency. This means that letters differ only in subtle differences of shading, a feature that we usually do not think of as shape. Observers would need to be able to utilize small intensity differences to distinguish between letters. To eliminate an alternative explanation (the smaller number of frequency components in the low-frequency bands), we conducted an informal experiment with a lower fundamental frequency. The fundamental frequency, which is outside the band, nevertheless determines the spacing of frequency components within the band. Reducing the fundamental frequency of the letter by one-half increases the number of frequency components in the band by a factor of 4. (A 256 × 256 sampling grid was used rather than 128 × 128.) These 4× more highly sampled stimuli were not more discriminable than the original stimuli. This suggests that the internal letter representation (template) that subjects bring with them to the experiment cannot utilize low-frequency information, even when it is abundantly available. Whether, with sufficient training, subjects could learn to use low spatial frequencies to make letter discriminations is an open question.

SUMMARY AND CONCLUSIONS

1. Visual discrimination of letters in noise, spatially filtered in 2-octave wide bands, is

independent of viewing distance (retinal frequency) but improves as spatial frequency increases.

2. The improvement in performance with increasing spatial frequency results mainly from an increase in the objective amount of information transmitted by the filters with increasing frequency (because filter bandwidth was proportional to center frequency) which is manifested as objectively less confusible stimuli in the higher bands.

3. The comparison of human performance with that of an estimated ideal discriminator demonstrates that humans achieve optimal discrimination (a remarkable 42% efficiency) when letters are defined by a 2-octave band of spatial frequencies centered at 1 cycle per letter height (mean frequency 1.5 c/letter). This high efficiency of discrimination is maintained over a 32:1 range of viewing distances.

4. Detection efficiency was invariant over a range of retinal spatial frequencies in which the contrast threshold for detection of sine gratings (the modulation transfer function, MTF) varies enormously. The independence of detection performance and retinal size held for all frequency bands.

5. A part of the loss of human efficiency in discrimination as spatial frequency exceeded 1 c/object height may have been due to the subjects' inability to identify, to selectively attend, and to utilize the smaller fraction of information-rich pixels in the higher frequency images.

6. Finally, it is important to note that without the comparison to the ideal observer, we would not have been able to understand the components of human performance in the different frequency bands.

Acknowledgements—We acknowledge the large contribution of Charles Chubb to the formulation and solution of the ideal discriminator. We thank Michael S. Landy for helpful comments and Robert Picardi for skillful technical assistance. The project was supported by USAF, Life Sciences Directorate, Visual Information Processing Program, grants 85-0364 and 88-0140.

REFERENCES

Barlow, H. B. (1978). The efficiency of detecting changes of density in random dot patterns. *Vision Research, 18*, 637–650.

Barlow, H. M. (1980). The absolute efficiency of perceptual decisions. *Philosophical Transactions of the Royal Society, London B, 290*, 71–82.

Barlow, H. B. & Reeves, B. C. (1979). The versatility and absolute efficiency of detecting mirror symmetry in random dot displays. *Vision Research, 19*, 783–793.

Burgess, A. (1985). Visual signal detection—III. On Bayesian use of prior knowledge and cross correlation. *Journal of the Optical Society of America A, 2(9)*, 1498–1507.

Burgess, A. (1986). Induced internal noise in visual decision tasks. *Journal of the Optical Society of America A, 3*, 93.

Burt, P. J. & Adelson, E. H. (1983). The Laplacian pyramid as a compact image code. *IEEE Transactions on Communications, Com-34(4)*, 532–540.

Campbell, F. W. & Robson, J. G. (1968). Application of Fourier analysis to the visibility of gratings. *Journal of Physiology, London 197*, 551–566.

Carlson, C. R., Moeller, J. R. & Anderson, C. H. (1984). Visual illusions without low spatial frequencies. *Vision Research, 24*, 1407–1413.

Chubb, C., Sperling, G. & Parish, D. H. (1987). Designing psychophysical discrimination tasks for which ideal performance is computationally tractable. Unpublished manuscript, New York University, Human Information Processing Laboratory.

Davidson, M. L. (1968). Perturbation approach to spatial brightness interaction in human vision. *Journal of the Optical Society of America A, 58*, 1300–1309.

Fiorentini, A., Maffei, L. & Sandini, G. (1983). The role of high spatial frequencies in face perception. *Perception, 12*, 195–201.

Geisler, W. S. (1984). Physical limits of acuity and hyperacuity. *Journal of the Optical Society of America A, 1*, 775–782.

Geisler, W. S. (1989). Sequential ideal-observer analysis of visual discriminations. *Psychological Review, 21*, 267–314.

Ginsburg, A. P. (1971). Psychological correlates of a model of the human visual system. In *Proceedings of the National Aerospace Electronics Conference (NAECON)* (pp. 283–290). Ohio: IEEE Trans. Aerospace Electronic Systems.

Ginsburg, A. P. (1978). Visual information processing based on spatial filters constrained by biological data. Aerospace Medical Research Laboratory, 1(2), Dayton, Ohio.

Ginsburg, A. P. (1980). Specifying relevant spatial information for image evaluation and display designs: An explanation of how we see certain objects. *Proceedings of SID, 21*, 219–227.

Ginsberg, A. P. & Evans, P. W. (1979). Predicting visual illusions from filtered imaged based on biological data. *Journal of the Optical Society of America A, 69*, 1443.

Harmon, L. D. & Julesz, B. (1973). Masking in visual recognition: Effects of two-dimensional filtered noise. *Science, 180*, 1194–1197.

Janez, L. (1984). Visual grouping without low spatial frequencies. *Vision Research, 24*, 271–274.

Kersten, D. (1984). Spatial summation in visual noise. *Vision Research, 24*, 1977–1990.

Legge, G. E., Pelli, D. G., Rubin, G. S. & Schleske, M. M. (1985). Psychophysics of reading—I. Normal vision. *Vision Research, 25(2)*, 239–252.

Legge, G. E., Kersten, D. & Burgess, A. E. (1987). Contrast discrimination in noise. *Journal of the Optical Society of America A, 4(2)*, 391–404.

van Meeteren, A. & Barlow, H. B. (1981). The statistical efficiency for detecting sinusoidal modulation of average dot density in random figures. *Vision Research, 21*, 765–777.

van Nes, F. L. & Bouman, M. A. (1967). Spatial modulation transfer in the human eye. *Journal of the Optical Society of America, 57*, 401–406.

Norman, J. & Ehrlich, S. (1987). Spatial frequency filtering and target identification. *Vision Research, 27(1),* 97–96.

Parish, D. H. & Sperling, G. (1987a). Object spatial frequencies, retinal spatial frequencies, and the efficiency of letter discrimination. Mathematical Studies in Perception and Cognition, 87–8. New York University, Department of Psychology.

Parish, D. H. & Sperling, G. (1987b). Object spatial frequency, not retinal spatial frequency, determines identification efficiency. *Investigative Ophthalmology and Visual Science (ARVO Suppl.), 28(3),* 359.

Pavel, M., Sperling, G., Riedl, T. & Vanderbeek, A. (1987). The limits of visual communication: The effect of signal-to-noise ratio on the intelligibility of American sign language. *Journal of the Optical Society of America A, 4,* 2355–2365.

Pelli, D. G. (1981). Effects of visual noise. Ph.D. dissertation, University of Cambridge, England.

Shannon, C. E. & Weaver, W. (1949). *The mathematical theory of communication.* Urbana: University of Illinois Press.

Sperling, G. (1989). Three stages and two systems of visual processing. *Spatial Vision, 4 (Prazdny Memorial Issue),* 183–207.

Sperling, G. & Parish, D. H. (1985). Forest-in-the-Trees illusions. *Investigative Ophthalmology and Visual Science (ARVO Suppl.), 26,* 285.

Tanner, W. P. & Birdsall, T. G. (1958). Definitions of d' and n as psychophysical measures. *Journal of the Acoustical Society of America, 30,* 922–928.

van Tress, H. L. (1968). *Detection, estimation and modulation theory.* New York: Wiley.

Winer, B. J. (1971). *Statistical principles in experimental psychology.* New York: McGraw-Hill.

APPENDIX

Both sub-ideal and super-ideal discriminators must compute estimates of the likelihood that the stimulus $u_{k,b}$ was produced with template $t_{i,b}$ and noise n_b, where k is the letter used to generate the stimulus, i is an arbitrary letter, and b indexes spatial frequency band. Let x be an index on the pixels of the image: $1 \leqslant x \leqslant 8100$, for the 90×90 images of the experiments.

For the Monte Carlo simulations of the super-ideal discriminator, the unknown stimulus parameters, $\alpha_{i,b}$ and σ_N^2 are computed during stimulus construction, and their exact values are supplied to the discriminator *a priori*. The sub-ideal discriminator, however, must estimate these parameters from the data as follows.

Sub-Ideal Parameter Estimation

Recall that stimulus contrast is modulated for any pixel x in the image:

$$u_{k,b}[x] = \beta_{i,b}[t_{i,b}(x) + n_b(x)] + q_{i,b}(x). \quad \text{(A1)}$$

The scaling constant $\beta_{i,b}$ limits range of real values for each pixel, prior to quantization, to the open interval $(-0.5, 255.5)$; the addition of $q_{i,b}[x]$, called quantization noise, rounds off pixel values to integers.

For each bandpass filtered template $t_{i,b}$, we first compute the correlation $\rho_{k,i}$ of the template to the stimulus $u_{k,b}$:

$$\rho_{k,i} = \frac{\sum_x u_{k,b}(x) t_{i,b}(x)}{\left\{\sum_x [u_{k,b}(x)]^2\right\}^{1/2} \left\{\sum_x [t_{i,b}(x)]^2\right\}^{1/2}}. \quad \text{(A2)}$$

To compute the likelihood estimates for each template $t_{i,b}$, we must be able to reverse the effect of $\beta_{i,b}$. Thus we define $\alpha_{i,b} = 1/\beta_{i,b}$ and choose $\alpha_{i,b}$ so as to minimize the expression:

$$\sum_x [\alpha_{i,b} u_{k,b}(x)]^2 = \sum_x [\rho_{k,i} t_{i,b}(x)]^2. \quad \text{(A3)}$$

Solving for $\alpha_{i,b}$ gives us:

$$\alpha_{i,b} = \rho_{k,i} \left\{\frac{\sum_x [t_{i,b}(x)]^2}{\sum_x [u_{k,b}(x)]^2}\right\}^{1/2}. \quad \text{(A4)}$$

Finally we set:

$$\sigma_N^2 = \frac{1}{X} \sum_{x=1}^{X} [\alpha_{i,b} u_{k,b}(x) - t_{i,b}(x)]^2 \quad \text{(A5)}$$

where $X = 8100$, the number of pixels in the image.

Likelihood Estimation

With estimates of σ_N^2 and $\alpha_{i,b}$ for the sub-ideal discriminator, and the *a priori* values for the super-ideal discriminator, we can formulate a maximum likelihood estimator. By rearranging terms of equation (A1) and dividing both sides by β yields:

$$\frac{u_{k,b}(x)}{\beta} - t_{i,b}(x) = n_b(x) + \frac{q_{i,b}(x)}{\beta}. \quad \text{(A6)}$$

Substituting $\alpha_{i,b}$ for $1/\beta$, and by transposing into the frequency domain, denoted by upper-case letters and indexed by ω, we have:

$$\alpha_{i,b} U_{k,b}(\omega) - T_{i,b}(\omega) = N_b(\omega) + \alpha_{i,b} Q_{i,b}(\omega). \quad \text{(A7)}$$

Note that the *left side of equation* (A7) is simply a difference image between the stimulus $U_{k,b}(\omega)$ and the template $T_{i,b}(\omega)$. This difference is exactly equal to the sum of the luminance and quantization noise only when the correct template is chosen ($i = k$). When the incorrect template is chosen ($i \neq k$) the right hand side of equation (A7) is equal to the sum of the noise sources plus some residue that is equal to $T_{k,b}(\omega) - T_{i,b}(\omega)$. Under the assumption that quantization noise can be modeled as independent additive noise in the frequency domain, the density Λ of the joint realization of the right-hand side of equation (A7) is given by:

$$\Lambda = \prod_\omega \frac{X}{\pi[\sigma_Q^2 \alpha_{i,b}^2 + \sigma_N^2 |F_b(\omega)|]^2} \times \exp\left[\frac{-x|\alpha_{i,b} U_{k,b}(\omega) - T_{i,b}(\omega)|^2}{\alpha_{i,b}^2 \sigma_Q^2 + \sigma_N^2 |F_b(\omega)|^2}\right] \quad \text{(A8)}$$

where $F_b(\omega)$ is simply the kernel of filter b, in the frequency domain. Dropping the multiplicative term in equation (A8), which does not depend on the template T, and taking logs, the ideal discriminator chooses the template that minimizes:

$$\sum_\omega \frac{X|\alpha_{i,b} U_{k,b}(\omega) - T_{i,b}(\omega)|^2}{\alpha_{i,b}^2 \sigma_Q^2 + \sigma_N^2 |F_b(\omega)|^2}. \quad \text{(A9)}$$

Finally, it is more convenient to compute the power of the quantization noise in the space domain (σ_q^2) than in the frequency domain (σ_Q^2); $\sigma_q^2 = \sigma_Q^2$. Spatial quantization noise, $q_{i,b}(x)$, is uniformly distributed on the interval $[-0.5, 0.5)$, so that σ_q^2 is computed as:

$$\int_{-0.5}^{0.5} x^2 \, dx \quad \text{(A10)}$$

and is equal to 1/12.

In D. E. Meyer and S. Kornblum (Eds.), *Attention and Performance XIV: Synergies in Experimental Psychology, Artificial Intelligence, and Cognitive Neuroscience - A Silver Jubilee* Cambridge, MA: MIT Press, 1992. (In press.)

Using Repetition Detection to Define and Localize the Processes of Selective Attention

George Sperling and Stephen A. Wurst

12.1 Introduction

Overview

In our repetition detection task, subjects search a rapid sequence of 30 frames for a stimulus that is repeated within four frames. Successful detection implies that a match occurs between an incoming item and a recent item retained in short-term visual repetition memory (STVRM).

When subjects attempted to attend selectively to subsets of items based on gross physical differences (such as color or size), they were unable to exclude the unattended items from STVRM. Frequently, there was better STVRM for unattended than for partially or fully attended items: this indicates that attentional filtering occurs more centrally than STVRM. That is, when to-be-attended items are defined only by their physical features, and not by space or time, there is no perceptual filtering prior to STVRM.

To explain such paradoxical results, we propose that selective attention attaches an attentional tag A+ to an item. A+ functions like a stimulus feature. All items are entered equally into memory. A+ items preferentially match other A+ items, and unattended A- items preferentially match A- items. In contrast to repetition detection, feature-based attentional selection does occur in partial report from visual streams. Therefore, if there is a single processing path, the locus of attentional selection is constrained to lie between the loci of repetition detection and of feature-based selection for partial reports.

Background: Early versus Late Selective Filtering

Theories of selective attention postulate that the human information processing system is limited in its capacity and that attention serves to select information to be processed from other, competing information (e.g., Broadbent 1958; Deutsch & Deutsch 1963; Norman 1968). Indeed, selective filtering of unattended information has been proposed as a mechanism in numerous visual processing tasks.

There is abundant evidence that selective attention can function as a mechanism to differentially filter information from different spatial locations (see reviews by Sperling & Dosher 1986; Sperling & Weichselgartner 1991). However, we find no convincing evidence that attention can function as a

mechanism for selecting information on the basis of physical features when items containing different constellations of features occur at the same location. Rather, the data are consistent with a theory that asserts that stimulus features serve only to guide spatial attention. That is, whenever selection appears on the basis of the physical features of visual stimuli (such as color, spatial frequency-filtering, size, etc.), these features serve to bring attention to a particular location, but the attentional filtering is on the basis of location rather than on the basis of feature. To test this theory, it is critical to present more information than can be successfully processed at a single location, and to observe whether, at this single location, attentional filtering is possible on the basis of physical features.

Selection from Streams

It is trivial to demonstrate that attentional filtering can occur within a given spatial location. Consider, for example, the following gedanken experiment. Subjects view a stream of alternating black and white digits on a gray background. Subjects are asked to compute the sum of the white digits and to ignore the black digits. Obviously, subjects can perform this task when the stream is slow enough, but this would not be profoundly revealing about selective attentional processes because we already know that selection can occur at a cognitive or a decision level of processing. The interesting questions about selective attention concern whether it can operate at an earlier sensory or perceptual level (reviewed in Sperling & Dosher 1986).

Search Procedures. A useful technique for studying attentional selection at a single location is to present a rapid stream of items at a location too rapidly to permit all items to processed perfectly. Attentional selection can then be used to determine which items are processed. There are a number of tasks that involve items that are presented in a rapid visual stream at a single location. For example, Sperling, et al. (1971) studied rapid visual search as a function of the number of locations in which streams of items were presented. However, the problem with search experiments is that, so far, no procedure has been developed to determine whether attentional selection (i.e., rejection of nontarget items) occurs at the perceptual or at the decision level of processing. Indeed, recent theories of selective filtering (Cave & Wolfe 1990; Duncan & Humphreys 1989; Pavel 1991; Wright & Main 1991; cf. Hoffman 1979) propose various cue-weighting algorithms to determine the sequence of attentional selections in visual search. Such weighting processes are typical of decision processes, although the algorithms themselves are neutral with regard to whether they operate at a perceptual or a decision level of processing.

Feature-based Partial Reports from Streams. Another task involving a stream is the selective recall of items according to their physical characteristics. The procedure involves the selection of items from a rapid stream according to whether or not the target items have a distinguishing characteristic such as a ring around them, or whether they are brighter than their neighbors. Subjects can extract single target items from a rapid stream (Intraub 1985; Weichselgartner & Sperling 1987), or even a short sequence of four targets (Weichselgartner 1984). In fact, such experiments are partial report experiments in which the the many items (from among which a few are selected for a partial report) are arrayed in time rather than in space as in the more usual procedure (Sperling 1960).

Feature-based Partial Reports from Spatial Arrays. In spatial arrays, subjects can select items for partial report that have a ring around them (Averbach & Sperling 1960) or items that merely are pointed at by a short bar marker--a minimal feature for selection. When subjects are required to report only items of a particular color from briefly exposed letter matrices, these partial reports are not much better than whole reports (von Wright 1968). Similarly, when subjects are required to report only digits from mixed arrays of letters and digits, subjects do not report more digits than when they must report both letters and digits (e.g., Sperling 1960). Both of these studies required subjects to extract both item-identity and location information from briefly exposed arrays. When subjects are required only to report the item identities and not locations, partial reports according to feature easily surpass whole reports (e.g., selecting solid from outline characters, Merikle 1980). Thus, with comparable response requirements, feature-cued partial reports are comparably successful in temporal streams and in spatial arrays.

Partial Reports according to Spatial or Purely Temporal (versus Featural) Cues. Originally partial reports were studied in spatial arrays, and the selection cue designated one of several rows of characters--purely spatial selection (e.g., Sperling 1960, 1963). With spatial cues, there is a large and consistent partial report advantage. When subjects must use a temporal cue to make a partial-report selection of four items from a rapid temporal stream, item selection appears to be based on a temporal window of attention (Sperling & Reeves 1980; Reeves & Sperling 1986, Weichselgartner & Sperling 1987). The subject's temporal window for selection from temporal streams is perfectly analogous to the spatial window for selection from spatial arrays (e.g., LaBerge & Brown 1989).

The Locus of Feature-based Attentional Selection. Partial-report paradigms primarily focus on the process whereby information is selected for inclusion in short-term memory. That feature-based attentional selection of information for partial reports can occur in streams and in arrays merely places the level of attentional selection below the level of short-term memory. This constraint is unremarkable. Therefore, it is search tasks that seem most often to have been called forth to resolve the issue of early versus late selection on the basis of physical features (recent examples include: Nakayama & Silverman 1986; Neisser 1967; Treisman 1977; Treisman 1986; Treisman & Gelade 1980; see Folk & Egeth 1989 for a review). Closely related issues are automatic versus controlled processing (Shiffrin & Schneider 1977), speeded classification (e.g., Felfoldy & Garner 1971; Garner 1978) and auditory selective attention (Swets 1984). The ambiguity of current search theories concerning the level of attentional selection was noted above. This is not the place for a review and critique of the many other approaches to these problems in the visual and auditory domains. Instead, we offer new variations of a repetition-detection task and new analyses that are particularly well suited to defining the locus of feature-based attentional selection (i.e., perceptual filtering according to physical properties).

Repetition Detection Paradigm

The repetition detection paradigm (Kaufman 1978; Wurst 1989; Sperling & Kaufman 1991) seems particularly well suited for the study of attentional selection based on physical features. In this paradigm (fig. 12.1) a stream of thirty digits is presented rapidly (typically, 9.1 digits per sec). Within this stream, every digit is repeated three times, but only one digit is repeated within four sequence positions (lag 4 or less);

all other digits are repeated with lags of nine or more. The subject is instructed to detect the recently repeated digit. Successful performance of this task obviously depends on the subject's ability to match incoming digits with previously presented digits in memory. Because all digits are repeated exactly three times within a list, only memory that discriminates short-lag repetitions from long-lag repetitions is useful for performing this task.

Figure 12.1

In previous research (Kaufman 1978; Sperling & Kaufman 1991), it was found that, at lag 1, repetition detection was typically better than 80% correct, and that by lag 4 it had dropped below 30 or 40 percent. Adding a noise field between successive frames (fig. 12.1) did not impair performance, even when the noise field was so intense that, if it were simultaneous with digit presentations, it would have rendered them illegible. This immunity to visual masking suggests a central memory locus for short-term visual repetition memory (STVRM), even at lag 1.

In another adaptation of the task (Kaufman 1978; Sperling & Kaufman 1991), it was found that using nonsense shapes as stimuli instead of digits yielded equivalent results. This suggests that STVRM is visual rather than verbal or semantic.

Wurst (1989) used dicoptic presentations to demonstrate that the locus of short-term visual repetition memory (STVRM) was after the locus of binocular combination. A particularly interesting finding in Wurst's dicoptic viewing procedure was that one eye was given priority over the other eye. Thus, a filtering of items by the eye of presentation may have been occurring even though items were presented alternately (never simultaneously) to the two eyes and though, in control conditions, monocular performance was the same for both eyes. The present study was undertaken to determine whether selection could occur by varying stimulus attributes other than the eye of presentation.

Plan of the Experiments

To investigate the role of attention in the short-term visual repetition memory task, as in the previous studies, digits are presented in the same spatial location while being viewed binocularly. Two levels of a dimension are employed (e.g., large and small sizes of digits), and digits alternate between the levels. We will call a level of a dimension a *feature*. For example, *small* and *large* are features within the size dimension. In this study, four stimulus dimensions that have typically been employed in attention research (e.g., Nakayama & Silverman 1986; Treisman 1982; Sagi 1988) size, angular orientation, spatial bandpass filtering, and contrast polarity (black-on-gray vs. white-on-gray), are examined separately. Additionally, we examine one feature pair (small-black versus large-white). Digits with a different feature (e.g., large and small size) are alternated at the same location. We determine the ability of subjects to attend selectively to items with one feature (or feature pair) while ignoring items with the other

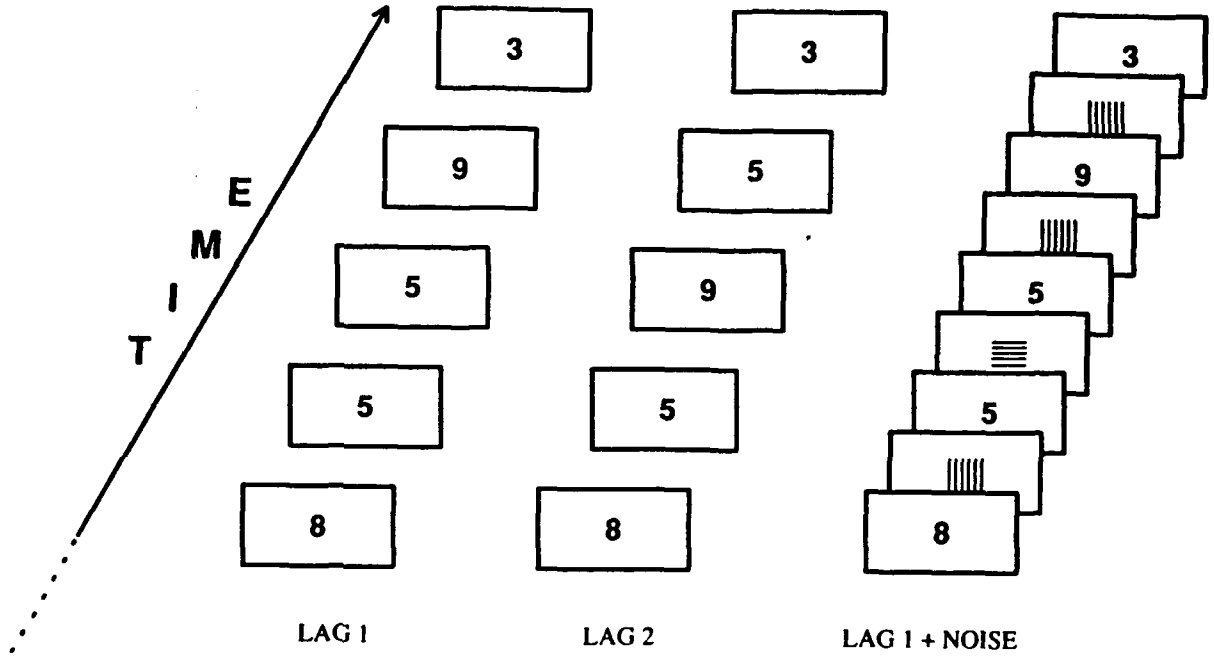


Figure 1. The repetition detection paradigm. The leftmost sequence (lag 1) represents five consecutive frames from the middle of a longer sequence of frames. The target repetition is the digit five. The middle sequence illustrates repetition of the digit 5 with lag 2. The rightmost sequence illustrates Kaufman's (1978) noise condition with lag 1. A grid of randomly chosen vertical or horizontal lines is interposed between each digit frame; repetition detection performance was unimpaired.

feature (or feature pair).

12.2 Method

In experiment 1, four stimulus dimensions are examined individually: size, orientation, spatial bandpass filtering, and contrast polarity. In experiment 2, two features are varied simultaneously to determine whether enhancing the difference between alternating stimuli would enhance attentional selection. The procedures of experiments 1 and 2 are quite similar so, although they were conducted sequentially, we consider them together throughout this chapter.

A stimulus sequence consists of 30 consecutive digits. A position in the sequence is called a *frame*; thus we say the i-th digit occurs in the i-th frame. Stimuli in a sequence alternately exhibit one level A of a dimension on odd numbered frames, and the other level B of the same dimension on even numbered frames. We call such as sequence $\frac{1}{2}A+\frac{1}{2}B$. If subjects were completely successful in selectively filtering out unattended B stimuli on the even numbered frames, detection of the repetitions of the attended-to-feature in a $\frac{1}{2}A+\frac{1}{2}B$ sequence would be similar to a control condition ($\frac{1}{2}A$) in which the even numbered frames were simply blank. If the selection were totally unsuccessful, for example, if the features were indiscriminable, then the alternating feature sequence should be as difficult as a same-feature sequence (A). Consider performance in the two control conditions $\frac{1}{2}A$ and A. The point between these two performances where performance with $\frac{1}{2}A+\frac{1}{2}B$ falls indicates the success of attentional filtering. This is the broad plan of the experiments. Additional complications will become apparent as the story unfolds.

Stimulus Generation

Frames. The repetition detection procedure (Kaufman 1978; Sperling & Kaufman 1991), was used in this experiment. Each trial consisted of a stream of 30 digits displayed on a video monitor. A digit was painted three times (three refreshes), followed by six refreshes of a blank, gray screen, all at 60 refreshes per second. The sequence of nine refreshes (digit plus subsequent blank screen) is called a *frame*. The frame duration is 150 msec; equivalently, the digit-to-digit stimulus onset asynchrony (SOA) is 150 msec. A digit sequence was composed of thirty frames: the 10 digits, each presented three times.

Lag. To distinguish the different types of repetitions that occur, we use the term *lag*. When a digit occurs in frame i of the sequence, and then again in frame j, $1 \le i < j \le 30$, the digit is defined as being repeated with lag $i - j$ (see fig. 12.1). Only the target digit was repeated within a lag of 4 or less; all other repetitions of the digits were separated by 8 or more intervening digits (lag ≥ 9). To generate a stimulus sequence, the first digit is chosen randomly. Subsequently, at any point in sequence generation, the requirement that no digit be repeated with lag ≤ 8 restricts the number of digits eligible to be chosen. At each point, the new digit was chosen with equal probability from among the eligible digits. The

critical repetition was embedded randomly in the sequence, with the restriction that the first member of the repetition pair occur between sequence positions 11 and 20. Each sequence was generated by a new random draw.

Figure 12.2

Figure 12.2a shows a typical sequence of thirty digits. Figure 12.2b shows the expected distribution of lags in such a sequence. A single lag of 1, 2, 3, or 4 represents the to-be-detected repetition--the signal. All the other repetitions have lag ≥9 and represent the noise. The distribution of noise lags is approximately exponential; it is truncated because repetition lags greater than 21 are impossible. While the actual noise distribution of lags is well defined, the *effective* noise distribution depends somewhat on how precisely, in such a rapid sequence, subjects can use their knowledge of constraints on the frames in which repeated pairs are permitted to occur (see below).

Procedures. Subjects were instructed to detect the repetition of lag 4 or less, and not to respond to any of the other stimuli. No masking stimuli were interleaved between the digits. All digits were presented in the same spatial location, centered on the CRT screen.

A trial began with a centrally located fixation square. When the subject was ready to begin the trial, the subject pressed any key on the computer keyboard. After a repetition was detected, the subject pressed the RETURN key as quickly as possible. After the end of the sequence, a message was presented on the monitor that cued the subject to enter the repeated digit and to enter a confidence rating between 0 (very low confidence that the response was the repetition) and 4 (very high confidence that the response was the correct repetition). The actual repeated digit was then presented on the screen to give the subject complete accuracy feedback information. A message to press the RETURN key was displayed, following which, the fixation square for the next trial appeared.

Stimulus Sets

Subjects viewed all stimuli at a distance of 93 cm. The square fixation box was a 2.46 x 2.46 deg visual angle. The digits 0 to 9 were used in the Times-Roman font. The background of all displays and blank intervals was set at 50 cd/m2^2. Unless otherwise specified, digits were white on gray, with a digit height of 0.74 deg.

Figure 12.3

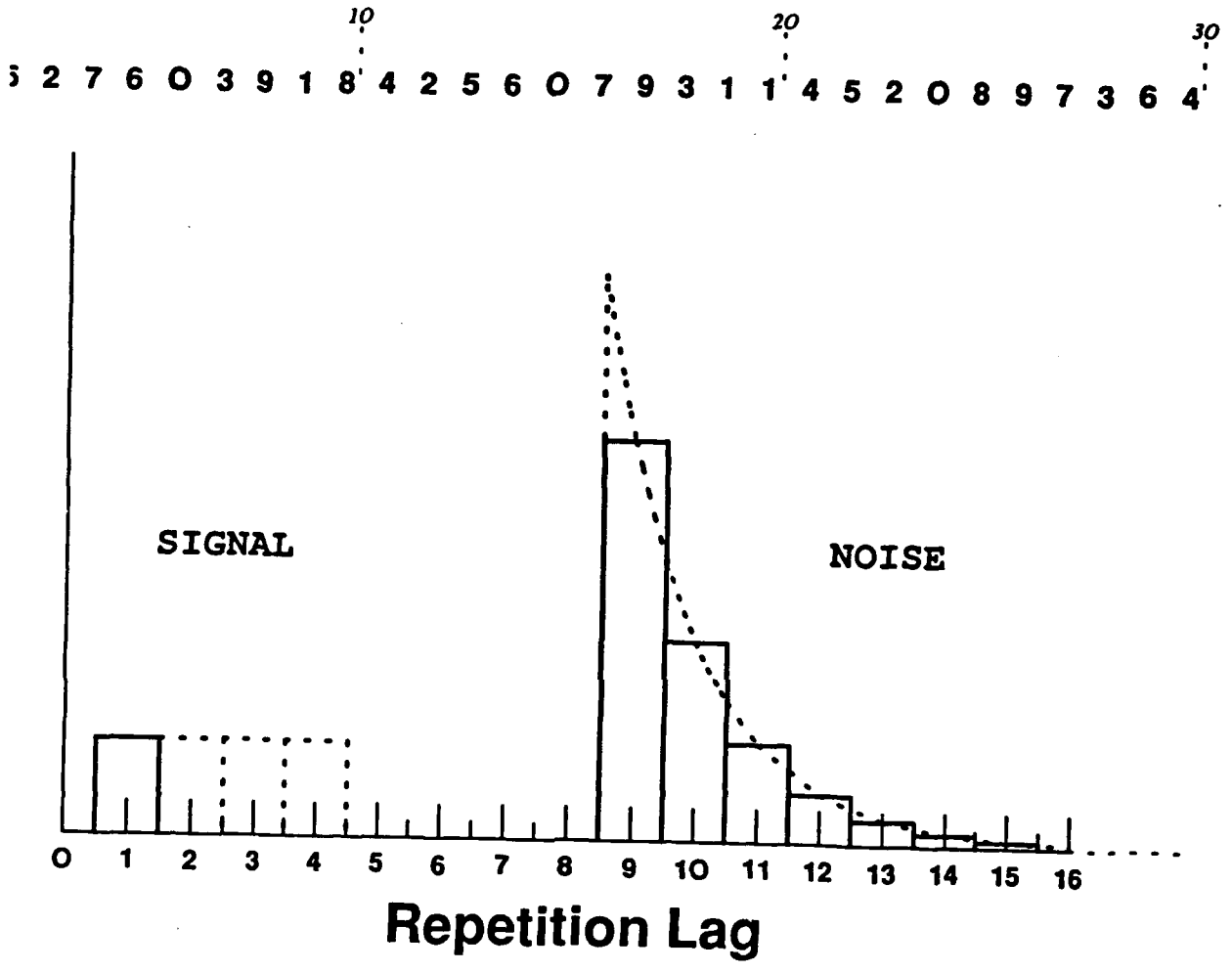


Figure 2. Top: A stimulus sequence in the repetition detection experiment. Bottom: The expected frequency distribution of signal (target) and noise repetitions. SIGNAL indicates that, on each trial, there is exactly one signal repetition, its lag is either 1,2,3 or 4. Nontarget digits are constrained to repeat only with lags of 9 or more (NOISE repetitions). The numbers 10 and 20 (top) demark the the middle ten positions of the sequence within which the initial element of the target repetition is constrained to occur. These two constraints determine the expected frequency distribution of noise repetitions, indicated as NOISE.

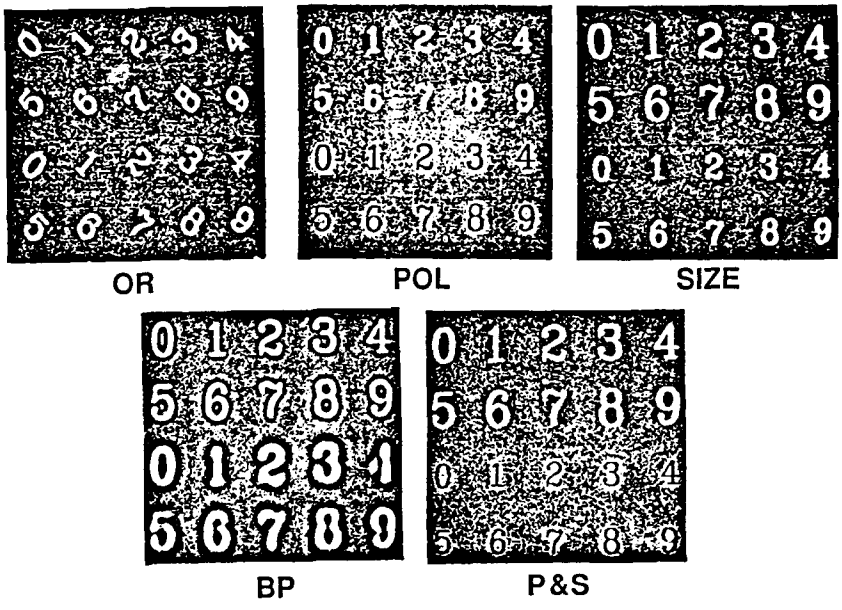


Figure 3. Stimuli used in the experiments. In each panel, the top 10 digits are the type A stimulus of the indicated dimensions (orientation, polarity, size, bandpass, polarity & size). The bottom 10 digits are the type B stimuli.

Four stimulus dimensions were investigated separately in experiment 1. There were two levels (feature values) for each of the four dimensions. The stimulus sets are shown in figure 12.3. The four dimensions (and the two feature values of each, A and B, respectively were

1. *size* (large, 0.74 deg visual angle versus small, 0.49 deg visual angle);
2. *orientation* (slanted 45 degrees up-to-the-left versus slanted 45 degrees right);
3. *contrast polarity* (white digits on gray background versus black digits on gray). The luminance level of the white digits was 101.50 cd/m^2, and the luminance level of the black digits was 0.40 cd/m^2 against a background of 50 cd/m^2.
4. (4) *bandpass filter* (high spatial bandpass versus low bandpass filtered). The mean luminance level for all bandpassed filtered stimuli was 50 cd/m^2. The high bandpass digits had a mean 2D frequency of 5.77 cycles per letter height, and the low bandpass digits had a frequency of 2.92 cycles per letter height. (See Parish & Sperling, 1991, for a description of the filters.)
5. *Polarity and Size.* These stimuli were used in Experiment 2. Large white digits represented feature type A; small black digits were type B. All were presented against the gray background. (Large, small, light, dark, gray were as defined above.)

Figure 12.4

Blocks of Trials

Figure 12.4 illustrates the design of experimental and control stimulus sequences and presents examples. A block of trials contained only one of the five stimulus transformations (fig. 12.3). Three experimental blocks all were of type ($\frac{1}{2}A+\frac{1}{2}B$) in which streams of strictly-alternating A,B stimulus features were presented. There were three kinds of experimental blocks for a given transformation that differed in the attentional conditions: attend to A, equal attention, attend to B). In addition to experimental blocks, which consisted of sequences that alternated two feature values (A and B), there were control blocks, which consisted of digits having the same feature value throughout.

Experimental blocks contained 100 trials, and control blocks contained 150 trials. Each of the trials was classified according to lag 1, 2, 3, or 4. In the experimental ($\frac{1}{2}A+\frac{1}{2}B$) blocks, trials were classified according to whether the the repetition pair was aa, ab, ba, or bb. [We use A and B to denote features or streams that contain the features (e.g., A = large and B = small). We use a, b, respectively, to denote target digits--members of the repetition pair--that contain feature A and B, respectively.]

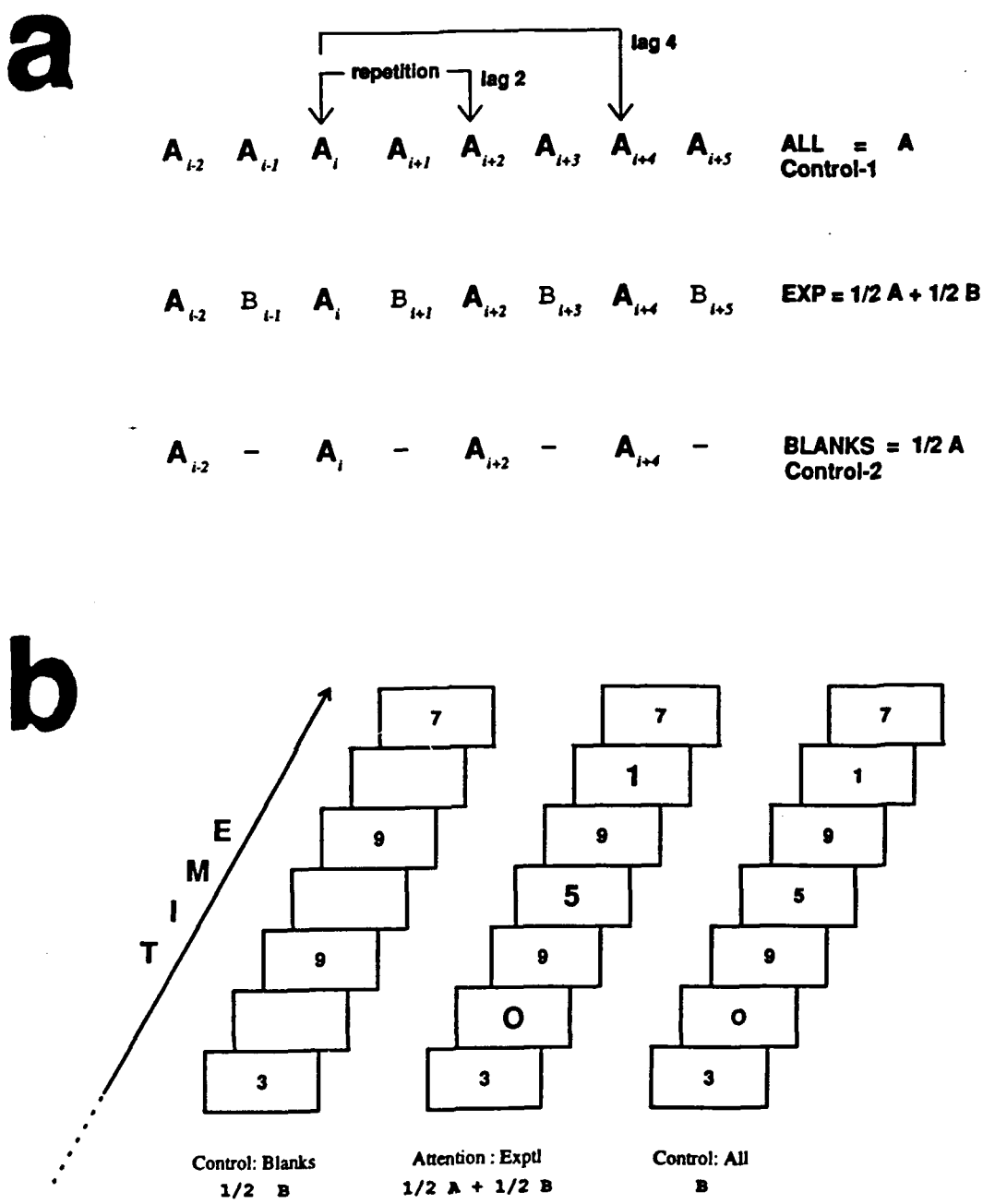


Figure 4. Experimental and control presentation sequences used to estimate the effectiveness of attentional filtering. (a) The middle row indicates the experimental condition, an alternating sequence of type *A* and type *B* stimuli, designated as $\frac{1}{2}A+\frac{1}{2}B$. If the subject could not discriminate the features that distinguished the type *A* and type *B* stimuli, the subject would perform equivalently in the $\frac{1}{2}A+\frac{1}{2}B$ and in to the "All" control, which consists entirely of *A* stimuli, designated simply as *A*. On the other hand, if the subject were able to perfectly ignore the unattended *B* feature in the $\frac{1}{2}A+\frac{1}{2}B$ stream, experimental performance would be equivalent to the "Blanks" control, designated as $\frac{1}{2}A$. This would be true for repetitions at lag 2 and at lag 4 (indicated above). (b) Graphical illustration of the three types of displays. The dimension is size. Type *A* stimuli are large, type *B* are small; the example illustrates bb detections.

Attention Conditions

The three experimental blocks are distinguished by the attentional instructions, the probability of the different types of repetitions presented, and the payoffs for correct responses. For the attend-A experimental block the subject was instructed to devote 80% of attention to feature *A* (e.g., large) and 20% to feature *B* (e.g., small); for the attend-B experimental block, the subject was instructed to devote 80% of attention to feature *B* (e.g., small) and 20% to feature *A* (e.g., large). In equal attention experimental blocks, the subject was instructed to devote 50% of attention to feature *A* and 50% to feature B. The probabilities of different trial types for the attend-A, attend-B, and attend-equal blocks are shown in table 12.1. Note that when attending to feature *A*, 70% of the trials in the selective attention blocks are pure (*a*, *a*) repetitions (35% at lag 2, 35 percent at lag 4). The remaining trials consist of mixed repetitions at lags 1 and 3, (*a*, *b*) 10%, (*b*, *a*) 10%, and of pure unattended-feature repetitions at lags 2 and 4, (*b*, *b*) 10%. The converse holds when attending to feature *B*.

Table 12.1

The attention instructions served only to define the initial conditions for the subjects. The steady-state behavior of subjects was controlled by carefully defined rewards to enforce the attention conditions. For every stimulus repetition in the attended-to stream that the subject detected correctly (that is, an *aa* or *bb* pair), the subject received 5 points. The subject received only 1 point for detecting repetitions in the unattended stream, and zero points for for the mixed ab and ba repetitions. The two paid subjects were paid 1 cent per point (in addition to their usual hourly wage for participation). The maximum expected payoff per trial for detecting targets with the attended feature is their probability of occurrence (0.7, table 12.1) times their value (5 cents), a net of 3.5 cents. The maximum expected earnings from detecting targets with the unattended feature is 0.1 × 1 cent = 0.1 cent. Thus, the expected value of detecting repetitions with the attended-to feature was 35 times greater than the value of unattended-feature repetitions. The 35:1 attended/unattended ratio of maximum possible earnings exerted a potent control over attention, although some of the effects of attention were unanticipated.

100% - 0% Attention Conditions. Even the extreme divided attention conditions (nominally 80% to 20%) involve divided attention because, when the subject notices repetitions involving the unattended feature, they are reported. Why not include experimental conditions in which the subjects are told to give 100% (rather than 80%) of their attention to the attended feature, are told to give 0% (rather than 20%) of their attention to the attended feature, and are paid only for detecting attended-feature repetitions? In previous research, Sperling & Melchner (1978a, 1978b) compared 100% - 0% attention to a range of divided attention conditions similar to the nominal 80% - 20% range used here. Sperling and Melchner's attentional manipulation involved only instructions; in contrast to the present study, their instructions were unenhanced by differential probabilities of occurrence of or by differential rewards for detecting attended

Table 1. Probability of Each Condition Within Each Block of Trials.

Blocks of Alternating-Feature Sequences (AB)

	Attend A		Attend B		Equal Attn.	
Target=	A	B	A	B	A	B
Lag 1*	.05	.05	.05	.05	.07	.07
Lag 2	.35	.05	.05	.35	.18	.18
Lag 3*	.05	.05	.05	.05	.07	.07
Lag 4	.35	.05	.05	.35	.18	.18

Blocks of Single-Feature Sequences

	Feature A		Feature B	
Stim.=	AA	A-	BB	B-
Lag 1	.167	--	.167	--
Lag 2	.167	.167	.167	.167
Lag 3	.167	--	.167	--
Lag 4	.167	.167	.167	.167

* Mixed-feature repetition pairs; "Target" indicates the feature of the first element of the pair.

targets. Nevertheless, in one-third of their cases, Sperling & Melchner's (1978b) divided-attention conditions spanned a range of performances that was fully as great as the extremes of the 100% - 0% control conditions, and their remaining divided-attention cases spanned most of the 100% - 0% performance range. Thus, while 100% - 0% conditions might (or might not) slightly expand the range of performances observed here, the added conditions would not be expected to produce any qualitatively different data.

Controls (A, B, $\frac{1}{2}A$, $\frac{1}{2}B$). Control blocks were run for each feature, as indicated in fig. 12.4 and in table 12.1. In the control-ALL trials (*A* and *B*), all thirty digits have the same feature value, and lags 1, 2, 3, and 4 occur equally often. Control-ALL trials were interleaved with control-BLANK trials ($\frac{1}{2}A$ and $\frac{1}{2}B$) in which every other digit in the sequence was replaced by enough blank frames to permit the remaining digits to retain their precise temporal positions in the sequence. Therefore, for control-BLANKS, only 15 digits were presented, and repetitions only occurred at what, in the ALL sequence, would have been called lags 2 and 4 (since blanks occurred at lags 1 and 3). As indicated in table 12.1, the six control conditions with feature *A* (or feature *B*) had an equal probability of occurring (i.e., 25 trials for each condition in the control blocks).

Altogether, there were 36 different kinds of trials for each of the five stimulus transformations (fig. 12.3). There were 24 experimental conditions: 4 lags (1, 2, 3, 4) × 3 attentional instructions (80%, 50%, 20%) × two kinds of targets (*aa*, *bb* at lags 2, 4; *ab*, *ba* at lags 1, 3). And there were 12 control conditions: the control-ALL contained 4 lags (1, 2, 3, 4) × 2 features (*A*, *B*), whereas the control-BLANKS contained 2 lags (2, 4) × 2 features ($\frac{1}{2}A$, $\frac{1}{2}B$).

Four blocks of each experimental condition, and at least three blocks of each control condition, were conducted. This yielded a comparable number of trials for the major data points of interest, and at least 20 trials for each of the most infrequent conditions.

Apparatus

A desktop computer (an IBM-compatible AT personal computer) was used to present stimuli and collect subjects' responses. Stimuli were created with HIPS image-processing software (Landy, Cohen, & Sperling 1984a,b) and displayed using a software package (Runtime Library for Psychology Experiments, 1988) designed to drive an AT-Vista Videographics Adapter that produced black-and-white images on a NEC Multisync-Plus color monitor (with horizontal resolution of 960 dots, vertical resolution of 720 lines, and short persistence phosphors).

Subjects

One female and two male New York University graduate students with normal or corrected-to-normal vision participated in this research. Two of these subjects were paid for their participation, and the third was the experimenter. The three subjects were well practiced on the repetition detection procedure before the formal experiments began.

Results and Discussion

Because there are 36 data points for each of the five types of stimuli, presentation of the results is quite complex. We use three kinds of graphs. The first shows the attention conditions relative to the controls; the second shows attention-operating characteristics; and the third shows all 36 conditions on a single graph. We also table the fractional benefits conferred by feature mixing and by attentional manipulations. Because our observed data are quite at variance with what might be expected from such experiments, we begin with a display (fig. 12.5) that compares hypothetical expected data with actually observed data in a typical condition for a typical subject.

Figure 12.5

Definitions Illustrated with Hypothetical Data

Figure 12.5a shows the hypothetical data from a subject viewing the bandpass stimuli. The data in fig. 12.5a represent detections of *aa* (high-spatial frequency) repetitions in various contexts. The control conditions are $\frac{1}{2}A$ and A (pure high spatial frequency stimuli); the experimental conditions are $\frac{1}{2}A+\frac{1}{2}B$ (mixed high and low spatial frequency stimuli). Consider first the diagonal line connecting data from the control conditions $\frac{1}{2}A$ and A. The condition $\frac{1}{2}A$ represents a plausible upper bound on the attention conditions because it corresponds to what would be expected if the subject succeeded in ignoring B stimuli entirely: the B stimuli are processed equivalently to the blanks of $\frac{1}{2}A$. The control A represents a plausible lower bound on attention: the B stimuli are not discriminated from A stimuli. Thus, the projections of the diagonal line of fig. 12.5a (control conditions) on the vertical axis indicate plausible bounds on the range of attention effects (0.50 to 0.98).

In the experimental conditions $\frac{1}{2}A+\frac{1}{2}B$, full attention to feature A while ignoring B is represented by the the middle point on the diagonal line of fig. 12.5a. Full attention to A shows a benefit relative to the control-ALL-A condition but not as great a benefit as would occur if the B stimuli were replaced with blanks. The three points on the diagonal line of fig. 12.5a represent actual data of subject BL, bandpass lag 1, detection of high spatial-frequency digits.

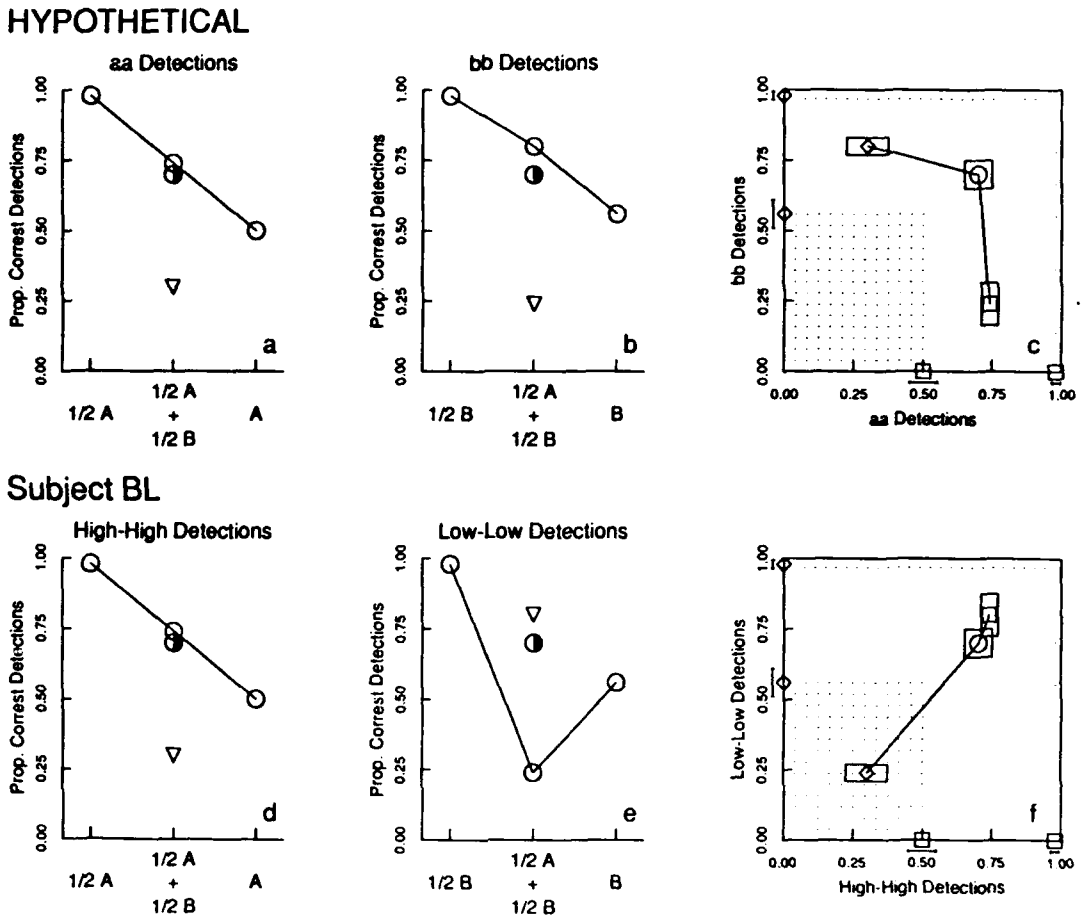


Figure 5. Hypothetical and actual results of an experiment on selective attention either to high (type *A*) or low (type *B*) bandpass filtered stimuli (see Fig. 3). (a, b, c) Hypothetical results. (a) The proportion correct in detecting *aa* (high-high) repetitions is shown as the ordinate for the three types of stimuli represented on the abscissa (see axis labels at bottom). The $\frac{1}{2}A + \frac{1}{2}B$ experimental displays serve three attention conditions: the data point for the attend-A condition is connected by lines to to the control conditions (which involve only *aa* repetitions); equal attention is the middle 1/2 tone point, and detecting aa while attending B is shown as the lowest point. (b) Data for detecting type *bb* (low-low) repetitions. Here, the point on the connecting line over $\frac{1}{2}A+\frac{1}{2}B$ represents attention directed to type *B* stimuli; the points underneath it represent equal attention and attend-A, respectively. (c) Attention operating characteristic (AOC) derived from the data of panels (a) and (b). The abscissa and ordinate, respectively, represent the proportion of correct *aa* and *bb* detections, respectively. The outer shaded area indicates performance better than a *blanks* control ($\frac{1}{2}A$, $\frac{1}{2}B$) for one or both of the two types of targets (*aa*, *bb*). The inner shaded area indicates performance worse than the corresponding *all* controls (A, B) for both *aa* and *bb* detections. The concave-down curve is the AOC derived from the $\frac{1}{2}A+\frac{1}{2}B$ stimulus with with the points representing, from left-to-right, attend-A, equal attention, and attend-B. The error bars indicate one standard error of the mean; the relative sizes of the errors derive from the inverse square root of the number of observations: small errors indicate many trials (attended features). (d, e, f) Real data from subject BL corresponding to the hypothetical data of (a, b, c). These data represent the most common type of result. (d) Coordinates and data as in (a). (e) Coordinates and data for control conditions are as same as (b) but the attentional data are in opposite order: The more attention the subject devotes to type *B* stimuli, the worse are *bb* detections. (f) The observed AOC for subject BL represents, from lower left to upper right, attend-low, equal attention, and attend-high bandpass stimuli. The AOC is perpendicular to the expected AOC.

Two hypothetical data points are shown in fig. 12.5a. The half-shaded point below the full attention point in fig. 12.5a indicates equal attention. We expect that equal attention in a mixed $\frac{1}{2}A+\frac{1}{2}B$ stream would yield better performance than in the control-ALL-*A* stream because mixing two features in the stream (instead of only one) makes the stimuli more discriminable. Attention to *B* stimuli is expected to lead to poor performance on *aa* repetitions (0.25), and this is shown indicated the triangle in fig. 12.5a.

In the hypothetical data, we expect complete symmetry between features *A* and *B*. So, fig. 12.5b, generated for detections of *bb* repetitions is basically the same as fig. 12.5a (except that the $\frac{1}{2}B$ and *B* points are based on real data for *bb* detections that are slightly different from the *aa* data of fig. 12.5a).

Attention Operating Characteristics (AOCs). AttentionOperatingCharacteristics(AOCS) The {1 over 2}*A*+{1 over 2}*B* points in figs. 12.5a and 12.5b generate the AOC (Kinchla 1980; Sperling & Melchner 1978b) of fig. 12.5c. The lower-right square of fig. 12.5c indicates joint performance on aa and bb repetitions when attention fig. 12. is directed to A. The rectangle around the square indicates one standard error of the mean in each dimension. The rectangle is extended in the *B* dimension because, in the attend-A condition, there are seven times more *aa* repetition trials than *bb* trials, and this increases the standard error of *bb* detections relative to *aa*. The circle in fig. 12.5c indicates equal-attention performance, and the diamond at the upper left end of the AOC indicates attend-B performance. Based on the hypothetical data of figs. 12.5a and 12.5b, the shape of the AOC is concave down, as expected.

Additionally, fig. 12.5c indicates two shaded areas that represent excluded performances. Regardless of the state of attention, we expect the subject to perform worse in any experimental $\frac{1}{2}A+\frac{1}{2}B$ condition than in the corresponding $\frac{1}{2}A$ or $\frac{1}{2}B$ control conditions. This excludes data from the shaded area in the outer rim of the AOC graph. And, we expect performance in $\frac{1}{2}A+\frac{1}{2}B$ to equal or exceed performance in the ALL-A and ALL-B control conditions. This excludes data from the lower-left rectangle of the AOC graph.

Definition: Fraction of Maximum Possible Benefits. The range between *A* and $\frac{1}{2}A$ defines the extent of possible benefits conferred by feature differentiation between *A* and *B* plus any additional benefits of selective attention. In fig. 12.5a, the maximum possible benefit extends from .50 and .98, a range of 0.48. The attend-A condition yields a fraction correct of .74, which is (.74 - .50)/(.98 - .50) = 0.50, exactly half of the possible benefit. For all the data points in fig. 12.5a, the hypothetical and the real data are the same. The equal attention condition yields a score of 0.70 as shown in fig. 12.5a and that attention to *B* would yield a score of .30. The fraction of possible benefit in the hypothetical equal-attention condition is (.70 - .50)/(.98 - .50) = 0.42. Equal attention involves only stimulus differentiation, not attention, so this fractional benefit of the alternating-feature stream is a *stimulus differentiation* benefit. Selective attention confers an additional benefit over and above the stimulus differentiation benefit.

In summary, the alternating-feature stream, $\frac{1}{2}A+\frac{1}{2}B$ confers two possible benefits: stimulus differentiation (in equal attention conditions) and attention-plus-stimulus differentiation (in selective attention

conditions). To estimate these benefits, it is useful to average over the two types of detections (*aa*, *bb*). That is, when alternating two features in the $\frac{1}{2}A+\frac{1}{2}B$ stream helps to differentiate stimuli, then detections of both *aa* and *bb* should be improved relative to the respective all-A and all-B controls. We define the *average achieved fraction of the maximum possible stimulus differentiation benefit*, Stim Benefit, as the improvement in equal-attention conditions (equal attention minus control-ALL) compared to the maximum possible range of improvement (control blanks minus control-ALL). To compute Stim Benefit, the following definitions are needed. Let $P(aa \mid \frac{1}{2}A+\frac{1}{2}B)_{Attn=A}$ be the probability of correct detections of *aa* repetitions given the $\frac{1}{2}A+\frac{1}{2}B$ stream with attention directed to the *A* feature. Let *A* indicate the ALL-A condition and $\frac{1}{2}A+$ indicates the *A* blanks control condition. Then,

$$Stim\ Benefit = \frac{1}{2}\left[\frac{P(aa \mid \frac{1}{2}A+\frac{1}{2}B)_{Attn=AB} - P(aa \mid A)}{P(aa \mid \frac{1}{2}A) - P(aa \mid A)}\right] + \frac{1}{2}\left[\frac{P(bb \mid \frac{1}{2}A+\frac{1}{2}B)_{Attn=AB} - P(bb \mid B)}{P(bb \mid \frac{1}{2}B) - P(bb \mid B)}\right] \quad (1)$$

Similarly, the *average achieved fraction of the maximum possible attention-plus-stimulus benefit*, abbreviated here simply to attention benefit (Attn Benefit), is

$$Attn\ Benefit = \frac{1}{2}\left[\frac{P(aa \mid \frac{1}{2}A+\frac{1}{2}B)_{Attn=A} - P(aa \mid A)}{P(aa \mid \frac{1}{2}A) - P(aa \mid A)}\right] + \frac{1}{2}\left[\frac{P(bb \mid \frac{1}{2}A+\frac{1}{2}B)_{Attn=B} - P(bb \mid B)}{P(bb \mid \frac{1}{2}B) - P(bb \mid B)}\right] \quad (2)$$

where *Attn=AB* denotes the equal-attention condition. For the hypothetical data, *aa* (fig. 12.5a) and *bb* (fig. 12.5b) detections were approximately symmetric, so the averages would be approximately the same as the *aa* values given in the earlier example. For the real data, Stim Benefit and Attn Benefit are tabulated in table 12.2 to be considered below.

AOCs for Real Data

Some Features Are Harmful to Attend. Figures 12.5d and 12.5e show the percent correct for *aa* and *bb* detections in the bandpass conditions for subject BL and lag 2. Detections of the high bandpass *aa* repetitions parallel the hypothetical expected data. Detections of the low bandpass *bb* repetitions in control conditions ALL-B and $\frac{1}{2}B$ are essentially equivalent to *aa* repetitions in control conditions ALL-A and $\frac{1}{2}A$. However attention conditions produce wildly different data. Selective attention to *B* results in the lowest proportion (0.24) of correct *bb* detections. The more attention is devoted to *B*, the worse are *bb* detections! Selective attention to feature *B* actually yields a *bb* hit rate of only 0.24 compared to a *bb* hit rate of 0.80 when attending to *A*. Selective attention to *B* produces different data than equal attention, but the direction of the difference is produce the same data that attending to *A* would have produced.

The inverse results for attending to low bandpass repetitions in fig. 12.5e combined with the normal results of high bandpass detections in fig. 12.5d yield an AOC in fig. 12.5f that is perpendicular to the expected AOC. Both subjects for both lags (2, 4) show this type of AOC in the bandpass condition.

Indeed, this AOC, which is perpendicular to the normal AOC, is the most commonly observed and perhaps prototypical AOC for these experiments.

Figure 12.6

Figure 12.6 shows all the AOCs from the experiments. The 22 AOCs represent five stimulus transformations, with lags 2 and 4, and all the subjects. In a few conditions, notably orientation, the effect of attention is quite small, but none of the 22 AOCs follows the normal concave down trajectory of the hypothetical data. Most AOCs have the prototypical shape that indicates attention to one feature (A) is helpful to both *aa* and *bb* detections whereas attention to the other feature (B) is harmful to both. The features that are harmful (versus helpful) to attend are small (versus large), low bandpass (versus high) and small-black (versus large-white).

Equal-Attention Can Be Harmful. In seven instances (all four polarity conditions; lag 4 orientation, subject BB; lag 4 polarity-and-size, subjects BB, SW), equal attention results in uniformly worse performance than either mode of selective attention.

Finally, even when attentional effects are quite small, as in the case of orientation, what effects there are tend to follow the same two patterns (harmful feature, harmful equal-attention) as have been noted for the other transformations. We momentarily defer explanations of these phenomena.

Benefits of Stimulus Differentiation and of Selective Attention

Under equal attention, the detection of *aa* repetitions was not significantly different from the detection of *bb* repetitions in any condition for any subject. Similarly, the *A* control conditions did not differ from the corresponding *B* control conditions. These results indicate that the *A* and *B* features were quite symmetric, and equivalent with respect to difficulty in the repetition detection task. Therefore, the benefit calculations, which are averaged over *aa* and *bb* detections, are representative of each type individually.

Table 12.2

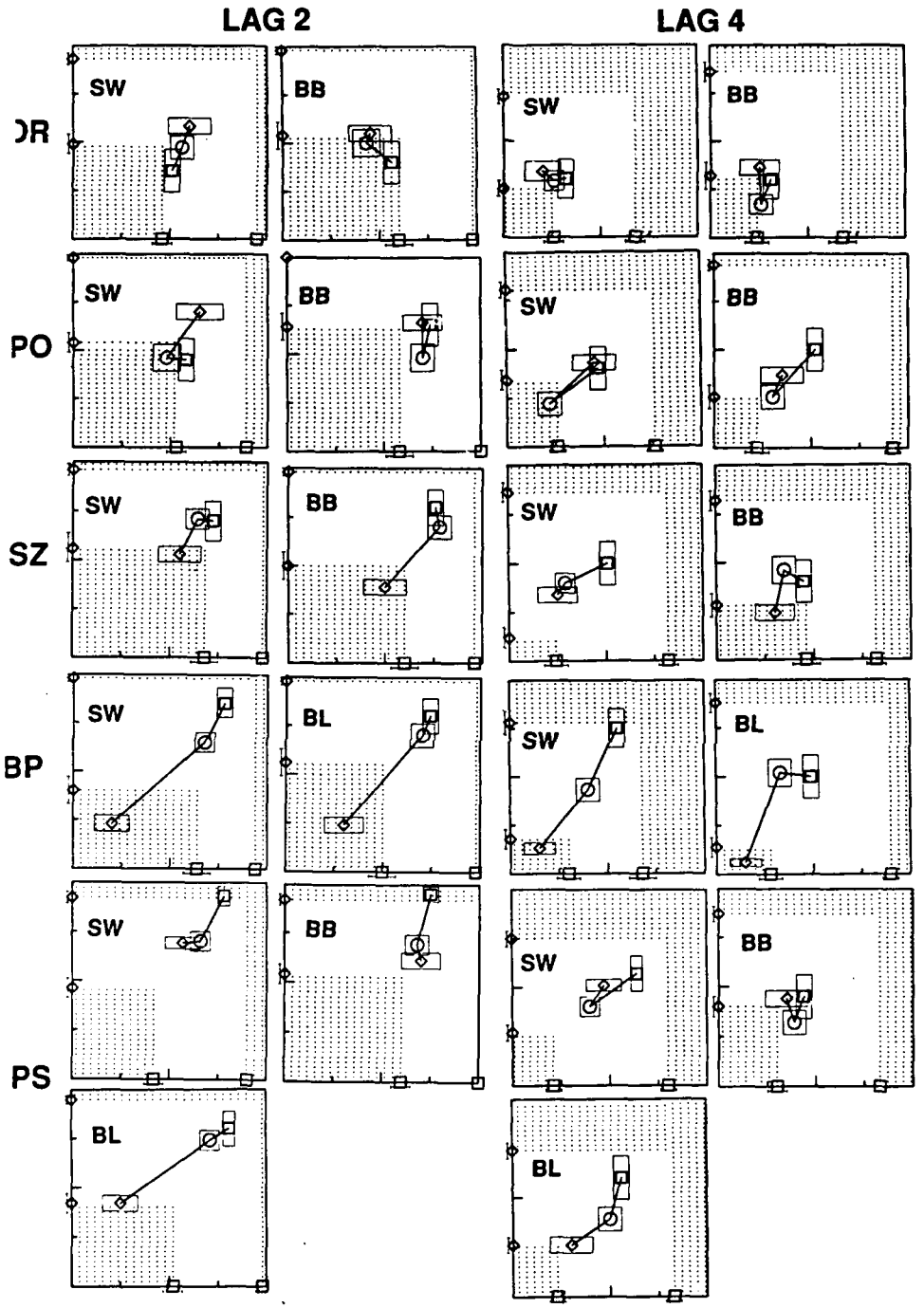


Figure 6. Attention Operating Characteristic (AOC) for all subjects and the five stimulus types of Fig. 3. The abscissa is the probability of detecting *aa* detections within each type stimulus transformation; the ordinate is the probability of bb detections. Symbols represent different attention conditions: squares = attend-A circles = equal attention, and diamonds = attend-B. Standard error bars are drawn around each point. Performances in A and $\frac{1}{2}A$ control conditions are shown on the abscissa, performances in B and $\frac{1}{2}B$ on the ordinate. The clear area defines the reasonable bounds on performance, given the control data. BB, BL, and SW indicate subjects.

Table 2. Fractions of the Maximum Possible Benefits Achieved in Equal Attention and in Selective Attention Conditions.

		STIMULUS BENEFITS[a]				ATTENTION BENEFITS[b]			
Subject:		SW	BB	BL	Mean	SW	BB	BL	Mean
ORIENT	lag2	0.08	-0.27	----	-0.09	0.15	-0.04	----	0.06
	lag4	0.04	-0.11	----	-0.04	0.17	0.12	----	0.14
	[2+4]c	0.06	-0.19	----	-0.07	0.16	0.04	----	0.10
POLARITY	lag2	-0.16	-0.08	----	-0.12	0.26	0.22	----	0.24
	lag4	-0.17	0.06	----	-0.06	0.31	0.30	----	0.30
	[2+4]	-0.17	-0.01	----	-0.09	0.29	0.26	----	0.27
SIZE	lag2	0.12	0.46	----	0.29	0.03	0.11	----	0.07
	lag4	0.23	0.01	----	0.12	0.38	-0.05	----	0.17
	[2+4]	0.18	0.24	----	0.21	0.21	0.03	----	0.12
BANDPASS	lag2	0.28	----	0.38	0.33	0.09	----	-0.13	-0.02
	lag4	0.34	----	0.30	0.32	0.29	----	0.12	0.20
	[2+4]	0.31	----	0.34	0.32	0.19	----	-0.01	0.09
POL-&-SZ	lag2	0.51	0.30	0.51	0.44	0.63	0.28	0.32	0.41
	lag4	0.30	-0.00	0.36	0.22	0.64	0.17	0.28	0.36
	[2+4]	0.40	0.15	0.43	0.33	0.63	0.23	0.30	0.39

a
Stimulus differentiation benefit measured in equal attention conditions, averaged over feature types A and B.

b
Includes stimulus differentiation plus selective attention benefits averaged over selective attention conditions. See text for computational details.

c
[2+4] indicates average of lags 2 and 4.

Stimulus Benefits. Table 12.2 shows benefits as calculated from equations 1 and 2. We consider first the stimulus benefits. These are determined under conditions of equal attention. If the subject were to violate the instructions and to have selectively attended one or the other feature of the stream in the equal attention $\frac{1}{2}A+\frac{1}{2}B$ condition, it would, in some cases, have improved performance and violate our assumption. However, the equality of the *aa* and *bb* equal-attention detections (and other internal consistencies in the data) suggest that the subjects did not adopt such a strategy.

The data from orientation and polarity illustrate minimal stimulus benefits. For example, consider the benefits of alternating two orientations in $\frac{1}{2}A+\frac{1}{2}B$ versus presenting a single orientation in A or B. For subject SW there is no benefit, for subject BB there is a slight loss in the $\frac{1}{2}A+\frac{1}{2}B$ conditions relative to the controls. With alternating contrast polarity, both subjects show a slight loss, rather than a benefit, in the $\frac{1}{2}A+\frac{1}{2}B$ condition. Alternating orientation or polarity stimulus features, in and of itself, is not helpful.

Both subjects show a small but clear benefit of selective attention to contrast polarity, and subject SW also shows a benefit of selective attention to orientation. This means that stimuli that are not well differentiated by their feature differences may nevertheless become differentiated by selective attention. In contrast to orientation and polarity, size, bandpass, and polarity-and-size show substantial benefits of stimulus differentiation. Apparently, these feature differences are recorded in STVRM and help to differentiate stimulus from noise repetitions.

Attention Benefits. Because the attention benefit is really an attention-plus-stimulus-differentiation benefit, we would expect it to exceed the stimulus benefit for all feature dimensions. Selective attention should enhance the stimulus differences. However, for the seven data sets that involve the significant feature differences (size, bandpass, polarity-and-size), only subject SW for polarity-and-size shows a consistently larger attention benefit than stimulus benefit. The other attention effects are quite variable and, in the case of bandpass (and some instances of size and of polarity and size), quite a bit smaller than the stimulus benefits.

The apparently negative incremental effect of attention is placed into context by noting that the attention benefit is computed to be the average of two states of attention. Size, bandpass, and polarity-and-size were the very conditions under which the paradoxical harmful effects of attention to the B dimensions were manifest. A harmful effect of attention to the attended B stream, counterintuitive as it seems, is fairly represented as a negative benefit that, in some instances, overwhelms the helpful effects of attention to attended A dimensions on the benefit computation. There is still a positive attention benefit in these particular conditions, but it is due to the residual stimulus benefit: the average incremental effect of attention is harmful to performance.

Additivity of Feature Differences. Polarity and size, by themselves, have certain stimulus and attention benefits (table 12.2). When polarity and size are combined in the polarity-and-size condition, the component benefits approximately add. Additivity is demonstrated by comparing the stimulus benefits for polarity *plus* the stimulus benefits for size with the stimulus benefits of polarity-and-size (averaged over subjects and lags). These values are -0.09 (polarity) + 0.21 (size) = 0.12, as compared to 0.33 (polarity-

and-size). The corresponding computations for attention benefits are b. efits are 0.27 + 0.12 = 0.39 (0.39). Within subjects and lags, the computations are a bit more noisy. It is interesting to note that additivity holds for attention benefits but fails for stimulus benefits, contrary to Dosher, Sperling, & Wurst (1986), who found perfect additivity of stimulus effects.

Figure 12.7

Consolidated Graphs of All Experimental and Control Conditions

Each panel of fig. 12.7 shows mean data for each of the 36 kinds of repetition detections for one subject and one set of features. Except for variances and tests of significance (Wurst 1989), these graphs represent the entire data of the experiments. The plan of fig. 12.7 is to indicate the data of control conditions by two sets of connected lines that form upper and lower reference bounds for four clusters of points that represent the data of the experimental conditions. We begin by making some general observations.

The Effects of Lag and SOA. The effects of lag on repetition-detection performance are indicated in fig. 12.7 by the sloping connecting lines that span lag 1 to lag 4. These sloping lines indicate performance in the control-ALL *A* and ALL *B* conditions. Performance with the control-ALL streams is at or above 90% at lag 1 for nearly all subjects and type of stimulus transformation. The two minor exceptions are the polarity-and-size conditions where, for two subjects, the control-ALL *B* slips down into the 80 - 90% range. By lag 4, performance drops into the 25% range, the largest drop occurring between lags 1 and 2. These data are completely consistent with earlier studies (Kaufman 1978; Sperling & Kaufman 1991; Wurst 1989).

The effect of SOA is derived from the sloping lines labeled *A*/2 that represent data for the control-BLANKS conditions ($\frac{1}{2}A$, $\frac{1}{2}B$), and which appear above lags 2 and 4. Performance in control-BLANKS is much better performance than the corresponding control-ALL (A, B) data. Alternatively, the control-BLANKS conditions with lags 2 and 4 might be described as lags 1 and 2 of a stream with a doubled SOA (stimulus onset asynchrony--the time from the onset of one digit to the next). However, the control-BLANKS is not quite equivalent to a slower sequence because it has only 15 instead of the 30 items that would be produced by simply slowing the stream. The combined manipulation of slowing and shortening the sequence produces (except for ceiling effects) much better performance for the comparable control-BLANKS than the control ALL conditions: control-BLANKS, lag 2, surpasses control-ALL, lag 1, and control-BLANKS, lag 4, surpasses control-ALL, lag 2.

The obvious interpretation of these data is that the main cause of the decline of performance with lag is retroactive interference (versus passive decay). Increasing the SOA increases the amount of time that the items must be retained but actually improves performance. (We know this also from unpublished observations in our laboratory in which sequence length was controlled.) Improved performance with

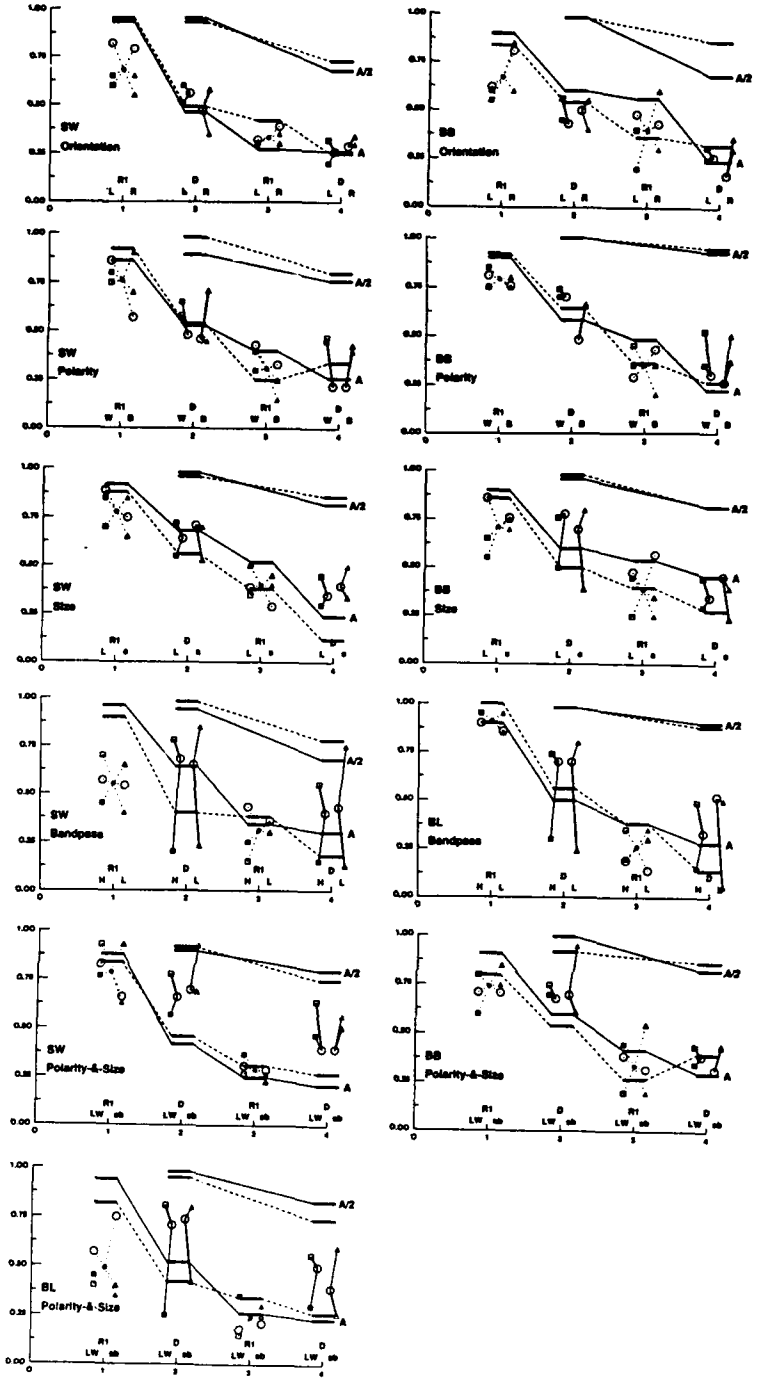


Figure 7. Data of all 36 trial types for each subject and type of stimulus transformation. In each panel, frame lag is plotted on the abscissa, and proportion of correct detections is plotted on the ordinate. Horizontal bars connected by continuous lines labeled A/2, A, represent control conditions $\frac{1}{2}A$ (blanks) and A (all). $\frac{1}{2}B$ and B conditions are indicated by bars connected by dashed lines (not labeled). The data points at each of the frame lags represent the different attention conditions and targets in $\frac{1}{2}A+\frac{1}{2}B$ stimuli. Frame lags 2 and 4 indicate *aa* and *bb* detections; frame lags 1 and 3 indicate mixed *ab* and *ba* detections. Open circles indicate equal attention. At frame lags 2 and 4, data points for the detection of *aa* repetitions are displaced to the left and detections of *bb* to the right, as indicated by dimension labels below) (D indicates "detection"). At frame lags 1 and 3, detections of *ab* repetitions are displaced to the left and detections of *ba* repetitions to the right. R1 indicates the first occurring feature in a mixed repetition pair, indicated by the dimension label below R1. Open symbols indicate detection of the attended feature (even lags) or detection of mixed repetition pairs in which the attended feature occurred first. Filled symbols indicate reports of unattended features or, in mixed pairs, that the attended feature occurred second. Reports of aa (bb) under different attention conditions are linked by lines, the heavier line indicated the attended feature. The asterisks at frame lags 1 and 3 indicate the means for the six mixed detection types.

increasing SOA suggests (1) the benefit of more time for encoding far outweighs the cost of passive decay within range of SOAs studied here, and (2) retroactive interference (versus passive decay) is the cause of diminished performance as a function of lag.

Repetition Blindness. The improvement of detection with shorter lags is different from another phenomenon discovered recently by using superficially similar procedures. "Repetition blindness" (Kanwisher 1987) is the reduced ability of subjects to report both occurrences of a repeated word embedded in a rapid sequence (approximately 4 to 9 per second) as contrasted with the reportability of two independent words. In contrast to the present research, reportability of both occurrences of the word increases with increasing lag. There are several differences between our repetition detection procedure and the procedure Kanwisher used. Repeated items are discriminated from unrepeated items in Kanwisher's studies rather than from other equally-often-repeated items, as in ours. The repetition blindness paradigm tests the tendency of subjects to report both occurrences of repeated items rather than their ability to discriminate repeated from unrepeated items. Moreover, repetition blindness experiments typically have used linguistic stimuli (words) in the stimulus sequence, in some instances varying the context in which these words were presented (Kanwisher 1987; Kanwisher & Potter 1989), and in other instances varying the case of the repetition without incurring a performance detriment (Marohn & Hochaus 1988). These procedural and stimulus differences suggest that repetition blindness and repetition detection paradigms may elicit different information-processing strategies and may reflect different levels of processing.

Equivalence of the Opposed Features within a Dimension. Feature equivalence has already been mentioned, but a glimpse at the control data shows that performance on the A and B control streams is essentially equivalent in all conditions. Indeed, there are several examples where the $\frac{1}{2}A$ performance is below $\frac{1}{2}B$ performance, and some where A performance is above B performance. None of these instances approaches statistical significance, considering that they are *post hoc* comparisons.

Feature equivalence means that differential attentional effects exhibited in the $\frac{1}{2}A+\frac{1}{2}B$ conditions are due to other factors than simple discriminability of the streams. Further, we note that attentional effects cannot be due to cross-stream masking in which an item from $\frac{1}{2}A$ masks one from $\frac{1}{2}B$. We refer again to the basic result that interposing noise fields (which are much more effective maskers) has minimal effects on performance (Kaufman 1978; Wurst, Sperling, & Dosher 1991). Furthermore, we have previously noted the equivalence of equal attention performance for *aa* and *bb* in $\frac{1}{2}A+\frac{1}{2}B$. In other words, at early levels of processing, the A and B items are equivalently discriminable.

Mixed Detections, ab and ba. In fig. 12.7, mixed detections are represented as clusters of points that lie above lags 1 and 3. Because of the strict feature alternation in the $\frac{1}{2}A+\frac{1}{2}B$ stream, only different-feature (mixed) repetitions can occur at lags 1 and 3. The probabilities of these repetitions were quite low, $P=0.1$ in the selective-attention conditions and $P=0.14$ in the equal-attention condition (table 12.1). At each of these lags, there are six mixed detection-types: three attentional states × two feature sequences (*ab*, *ba*). All six detection types are illustrated in fig. 12.7 for each stimulus transformation, subject, and lag.

Because a mixed repetition involves a feature difference, we expect mixed repetitions to be more poorly detected than same-feature repetitions in all conditions. The mean of all six (fig. 12.7) mixed repetition types for lag 1, is below the level of same-feature repetitions in all 11 instances, and dramatically below the same-feature level in some instances. But the pattern is hard to discern. With lag-1 mixed repetitions in the bandpass transformation, subject BL almost equals his performance with same-feature repetitions, whereas subject SW's mixed detections are grossly impaired. In the polarity-and-size transformation, this data pattern is reversed for the two subjects.

At lag 3, however, mixed detections fall right at the mean of same-feature detections in about half of the 11 instances, and not far from the mean in the remainder. This suggests, as previously noted (Kaufman 1978; Sperling & Kaufman 1991), that physical features are most important at lag 1 and that more abstract memorial representations become more important at longer lags. Whether the decreased dependence on physical features with increasing lag represents two memory systems, or one memory system with different properties as a function of signal strength, is not resolvable here. Finally, although the theory that will be considered later suggests that there should be significant differences within six mixed detection types, we did not discover any strong consistent differences.

Main Effects of Feature Differentiation and of Selective Attention. The main attention effects involve same-feature detections (*aa*, *bb*) and are represented in the clusters of points above lags 2 and 4. Previously, in table 12.2, we observed that the benefits of feature differentiation appeared only for the size, bandpass, and polarity-and-size transformation types. Selective attention, on the other hand, impaired *average* performance for these transformations but enhanced performance with orientation and polarity transformations. In fig. 12.7, the details of these attentional phenomena are manifest. Equal attention is represented by two large circles, one for each kind of detection (*aa*, *bb*). Each equal-attention circle is connected to two "attentional" line segments that represent the directed attention conditions with the same target. In the top two rows of fig. 12.7 (orientation, polarity), the equal attention data occur roughly at the centroid of all the attentional conditions and fall more or less on the control-ALL lines. When selective attention is appropriately directed (heavy lines, open symbols) about half the data fall above the equal attention circles and above the control lines, thereby indicating a small benefit of selective attention. The remaining data show little effect of attention.

In the size, bandpass, and polarity-and-size conditions, the equal-attention centroids move clearly above the control-ALL lines, indicating a significant stimulus benefit. In many instances, selective attention data fall well above and far below the equal attention data, indicating large effects of selective attention. However, the appropriately and inappropriately directed attention have similar effects, the direction of the effect is determined by the direction of attention, independent of the type of target (*aa* or *bb*). Thus, on the average, even appropriately-directed selective attention is not more beneficial than equal attention. The thick-lined attention spokes (appropriately directed selective attention) that point both up and down in fig. 12.7 (instead of up only) from the equal-attention circles represent the misdirected AOCs of fig. 12.6.

12.4 Theory: Attention is Itself a Feature

Neural Network Theory of Short-Term Visual Repetition Memory (STVRM)

Wurst, Sperling, & Dosher (1991) outline a theory that accounts reasonably well for the basic properties of the data of their repetition detection experiments. Following an established literature (McClelland & Rumelhart 1988), it is assumed that items are represented in memory as vectors of (dozens of) features. Each item vector has a value for every feature that can be represented in short-term visual repetition memory (STVRM). The feature value may be present/absent or perhaps more finely graded. Features are assumed to represent properties of items such as, for example, "has a vertical stroke", "has an intersection", and so on. It is assumed that features are elementary visual components (rather than semantic components) because it was found that nonsense forms were remembered precisely as well in STVRM as meaningful alphanumeric stimuli (Kaufman 1978; Sperling & Kaufman, 1991). However, the present discussion does not depend on what the particular features are assumed to be.

Items are not retained perfectly in the model of STVRM because new items use the same limited memory capacity as old items. The net effect is that memory noise (item uncertainty) increases as new items are added to STVRM. If a subject had perfect temporal discrimination, only the 11-th to the 24-th items of the 30 item input stream would need to be processed. To determine whether a repetition has occurred in the stimulus stream, comparisons are made between immediately successive items entering STVRM. Additionally, comparisons are made between the incoming item and older items in memory. [It is worth noting that a memory that simply recorded features of items, and in which the incoming item was compared, featu..-by-feature, to the *sum* of all recorded items (e.g., Murdock 1982), would not deal adequately with attention tags because, once stored, the tags become a property of memory rather than remaining associated with items.] The outcome of all comparisons is a single number (the *familiarity strength value*) that combines the degree of similarity between the incoming item and the most-similar memory item, and the confidence in this similarity.

Given that a familiarity strength has been computed, it is still necessary to generate a response. Because there is only one to-be-reported repetition on each trial, generating a response requires finding the item that has the highest familiarity strength value. This is nontrivial because humans cannot store a strength value for a dozen incoming items at an input rate of 6.7 items per second and then pick the largest value at the end of a trial. An alternative algorithm that stores the strength value of the first item and updates it whenever a new item has a higher strength would be computationally too demanding. Instead, a slightly less than ideal strategy is proposed. A repetition is reported when one of the strength values exceeds a threshold criterion value. An optimal--or nearly optimal--criterion evolves during practice with feedback in the particular task. The model with these assumptions about memory and decision processes provides a reasonable account of both accuracy and confidence data from a wide range of repetition detection experiments (Wurst, Sperling, & Dosher 1991).

Attention Is a Feature: A+/A-

The critical addition to the just-described memory and decision theory is that, in addition to features that represent physical properties of the stimulus, we assume that selective attention to an item functions like a feature in STVRM. Thus, along with physically defined features such as "vertical stroke" there is an attentionally generated feature "attended," which can have the value A+ to indicate the item was attended, or A- to indicate that it was not attended.

Because, in STVRM, the A+/A- feature is assumed to function just like other stimulus features, A+ tagged items preferentially match each other. That is, attended items have the A+ feature in common, and this facilitates the detection of repetitions for A+ items. However, unattended A- items, also share a common A- feature, and detection of unattended A-, A- repetitions is facilitated just as much as attended A+, A+ repetitions. Normal performance operating characteristics that are observed in other tasks reflect mental operations subsequent to STVRM in which the attention feature is used, for example, to select items for memory (as in partial report) or for further processing (as in visual search).

The Argument against Early Perceptual Selection. If it were possible to selectively filter items so that they did not enter STVRM, then A-, A- repetitions would be undetected. However, the opposite result is observed: A-, A- repetitions are remembered as well as A+, A+ repetitions. The surprising result for about one-third of our conditions was that, when subjects gave equal attention to all dimensions, their repetition detection performance suffered relative to both the attended A+, A+ and the unattended A-, A- repetitions. In the remaining data, attention to the dominant feature enhanced performance for both *aa* and *bb* detections relative to equal attention or attention to the subordinate feature. The systematic superiority of repetition detection of unattended items (when attending to a dominant feature) over partially-attended and fully-attended items is an extremely robust finding, occurring for all subjects and several stimulus types. Unequivocal superiority of unattended over partially-attended items really is quite extraordinary, and certainly requires the unattended (as well as attended) items to enter memory. These results fall quite nicely out of the attention-is-a-feature theory.

Transformation-specific Biases in Attentional Tagging. There were two main classes of data in repetition-detection experiments: In type 1, equal-attention was harmful, and in type 2, selective attention to one of the two features was harmful. Type 1 data are characterized by the inferiority of both *aa* and *bb* detections under equal attention relative to both *aa* and *bb* under all conditions of selective attention. Type 2 is characterized by the superiority of both *aa* and *bb* detections under attention to feature *A* relative to equal attention or attention to feature *B*.

Type 1 data are explained by assuming that the subject can reliably attach the A+ tag to input items according to instruction. Since this tag is reliably attached, it is equally useful for matching attended pairs with attended pairs and unattended pairs with unattended pairs. Thus both attended and unattended items benefit from selective attention.

Obviously, in type 1 data (and therefore in this theory), selective attention does not operate by filtering out items from STVRM. Unattended items occupy just as much space in STVRM as do attended items, and they are just as available when the right question is asked. They could be filtered from subsequent processing because they have an A- tag: the subject knows that they were "unattended". But this would be at a higher level of processing, where discriminations are made among items that are already in memory.

Type 2 data are explained by assuming that the subject can attentionally tag all the items in the *A* stream when attending to *A* but cannot reliably tag all the items in the *B* stream when attending to *B*. Thus, in selective attention to *A*, the subject succeeds in doing what comes naturally: correctly tagging the *A* as attended and the *B* items as unattended. The subject gains the benefit of reliable attentional tagging, which facilitates repetition detections of both *aa* and *bb* pairs because each is tagged correctly.

Now, consider the problem of attempting to attend to *B* in the face of an inherent bias to attend to *A*. Suppose the subject fails to attend (attach A+ tags to) all the *B* items and inadvertently attends (attaches A+ tags to) some not-to-be-unattended *A* items. Such mistagging works symmetrically to the detriment of both *aa* and *bb* detections becauses mismatched attention tags disturb what otherwise would be matches.

Dominant Features. All three subjects detected the small-black digits better when attending to the large-white digits than when attending to the small-black digits. *Post hoc* this suggests that large-white is a "dominant" feature relative to small-black, in the same sense that figure is dominant relative to ground. There is an *a priori* bias to associate attention with large-white. When the subject attempts to attend to small-black, the rapid presentation rate does not allow enough time to overcome this bias. Similarly, large is dominant relative to small and high bandpass is dominant relative to low. However, in the simple polarity condition, the black-and-white streams seem to be more or less equivalent. This suggests that size is the operative factor in determining feature-dominance when it is involved together with polarity in S&P.

In the bandpass transformation, there is a one octave (2×) difference between stimulus bands. Because the two 2 dimensional spatial filters are scaled replicas of each other, there is 2×2 times more information in high bandpass images than in low bandpass images (in the sense of number of independent samples). In both cases, size and bandpass, the attention-attracting feature seems to be the one that would activate the larger number of neurons. This argument closely follows Milner's (1974) distinction between extrinsic and intrinsic attention (532), where extrinsic attention is determined by summed neural activity. With slower presentation rates, the attention-bias of features might be overcome by conscious effort (intrinsic attention). If so, extrinsic and intrinsic attention in the repetition detection procedure (and in simpler procedures) could be conjointly scaled (Krantz & Tversky 1971) together with the attention dominance of images and image transformations.

Mixed Pair Predictions. When attentional tagging is carried out correctly, mixed repetition pairs under selective attention should be impaired relative to equal-attention mixed pairs. Both kinds of mixed pairs should be impaired relative to homogeneous pairs because, in addition to attentional differences, there are real stimulus differences (e.g., Posner et al. 1969, name vs. physical identity matching) Under conditions in which attentional tagging is unreliable, the differences between equal-attention mixed, selective-attention mixed, and homogeneous repetitions are expected to diminish: mixed pairs benefit from mistagging while homogeneous pairs suffer.

These complex mixed-pair predictions are not borne out by the data. While there are some data sets in which equal-attention mixed pairs fare better than selective attention pairs (e.g., SW, orientation, lag 1), the effect is not reliable over all conditions. A particular problem is that, because the mixed trials represent a small fraction of the total trials, the reliability of available mixed-pair data is low.

In mixed-pair repetitions, feature differences overshadow attention differences. For example, the orientation feature has a large influence on shape and therefore greatly impairs mixed pair detections at lag 1 (but not lag 3). Size and polarity show significant but smaller mixed-pair deficits. While equal-attention mixed-pairs fare better in these transformations, the effect is not reliable. The variability of mixed pair detections is illustrated in the bandpass and P&S conditions. With bandpass stimuli, mixed pair deficits are enormous for subject SW and almost absent for subject BL, while the reverse is true with polarity-and-size stimuli.

Mixed Pairs as a Tool to Study Memory Representation. Variations in feature similarity between mixed pairs offer a means of studying their significance in STVRM. Surface feature properties such as orientation, polarity, size, and bandpass play a significant role for mixed detections at lag 1 but become insignificant at lag 3. Systematic manipulation of mixed-pair differences could, in principle, establish metrics for the dimensions that underlie the memory comparisons.

Relation to Other Paradigms, Early versus Late Selection

Spatio-temporal versus Featural Selection. In paradigms involving spatial selection, such as partial report (Orenstein & Holding 1987), or temporal selection, such as the attention-gating procedure (Sperling & Reeves 1980; Reeves & Sperling 1986), items outside the spatiotemporal window of attention are unavailable. These are examples of successful attentional selection. Reeves & Sperling (1986) give a fairly complete account of the temporal window of selection (the attention gate), and Sperling & Weichselgartner (1991) extend the account to the dynamics of spatiotemporal attentional selection. On the other hand, where attentional selection on the basis of featural properties has been asserted to occur, as in visual search, it has not been clear whether the features serve only to guide spatial or temporal attention or whether the features themselves serve as the basis for selection. The present data suggest that, within a location, early selection on the basis of gross physical features does not occur. Certainly featural properties can exert an influence on detection and other decisions, but our data suggest that featural effects occur at the level of decision making, not at the level of exclusion from processing.

The Processing Level of Feature-based Attentional Selection. In contrast to the present study, Intraub (1985) and Weichselgartner & Sperling (1987) apparently observed feature-based selection in character streams. Their procedure utilized a rapid stream of characters in which the target character was surrounded by an outline square or circle. Report of the target character can be nearly flawless even at stream rates exceeding 10 characters per second. Weichselgartner & Sperling (1987) also demonstrated that highlighting a character in a stream (making it brighter than its neighbors) allowed it to be extracted almost perfectly. This is in striking juxtaposition to the present experiment, in which white characters could not be selectively attended in the context of black characters and vice versa. Further, Weichselgartner (1984) demonstrated that when more than one character was cued (by an outline square or by highlighting), as many as four characters could be selected. Thus feature-based cueing, like spatial and temporal cueing, can yield multi-item selection.

There are two contrary results: feature-based selection from temporal streams for partial report and feature-based nonselection in the repetition-detection procedure. What do they imply about about attentional processing? Two possibilities are that partial report selection occurs at a higher level of processing or in a parallel processing path. We can exclude a single path in which partial report precedes STVRM because, if attentional selection could have occurred before STVRM, it would have been manifest by our attentional manipulations.

Suppose the processes underlying repetition detection and feature-selected partial reports share a single processing path. Then, the locus of attentional selection is rather closely constrained by these results to lie between STVRM and selection for partial-report. Considerations about the precise level of processing of STVRM become critical. We have already recounted the earlier observation (Kaufman 1978; Sperling & Kaufman 1991) that never-repeated nonsense visual shapes yield precisely the same data as do letters in the repetition detection paradigm. This implied that the STVRM was visual (not verbal, for example). Wurst (1989) showed that STVRM occurs after binocular combination, that it is blind to eye of origin, and that dividing a stream alternately between two spatial locations does not alter STVRM capacity within a location. Within STVRM itself, the situation is less clear because there are differences between lag 1 and longer-lag detections. Kaufman (1978) and Sperling & Kaufman (1991) had noted that manipulations such as varying character font or introducing spatial jitter selectively affected lag 1 detections relative to longer lags. Indeed, Sperling & Kaufman (1978) argued that this implied lag 1 utilized a different memory than did other lags.

With respect to feature selection for partial report, Weichselgartner & Sperling (1987) showed that using a simultaneous auditory click to cue an item from a rapid visual stream failed to uniquely select an item. Rather, the click caused reports typical of other cues that define an interval in time from which to select items, analogously to tonal cues in partial report that have been used to define an interval in space from which to select items. That is, the trial-to-trial distribution of items that subjects reported as simultaneous with the auditory click had a large temporal window that included some items before and many items after the cued item. Only visual cues served to cleanly extract the cued item. On the other hand, when the size of an outline square was varied, it was found that even squares many times larger than the target letter functioned as well for extraction as did close-fitting squares. The large differences in data from auditory cueing and from visual cueing of items in visual streams implies that visual cues for partial

report of visually presented items produce selection at a visual level. But the observed indifference to the size of a cueing square suggests that the memory representation is not a simple transformation of the stimulus.

This review of featural selection from visual streams by means of the repetition detection paradigm and the partial report paradigms suggest a quite similar visual basis for performance in both paradigms The observation that attentional selection is possible in partial reports but not in repetition-detection suggests that attentional selection occurs subsequent to STVRM and prior, or internal, to selection for partial report. The generally similar visual sensibilities of partial reports and repetition detections suggests that attentional selection for partial reports follows very closely, perhaps acting directly on, the substrate provided by STVRM.

12.5 Summary and Conclusions

Background: Detection of visual repetitions in a rapid stream of items depends on a short-term repetition memory (STVRM) that is indifferent to eye of origin and to interposed masking fields, and which functions as well for nonsense shapes as for digits. STVRM is visual, not verbal or semantic. It is governed by interference from new items; it does not suffer passive decay within the short interstimulus intervals under which it has been tested.

Our subjects' attempts to selectively attend to characters based on physical differences of orientation, contrast polarity, size, spatial bandpass filtering, and polarity-and-size combined yielded large (but unexpected) attentional effects. The most common result was that certain attentional states impaired repetition detection of both attended and unattended items. There was no evidence for attentional selection. Repetition detection of unattended items was not systematically impaired relative to partially or fully attended items. Although some of the feature differences produced large stimulus benefits that aided detection in alternating-feature streams, they never produced attention benefits that consistently exceeded the stimulus benefits.

The paradoxical results were explained by assuming that attention itself functions like a stimulus feature, with values A+/A- to represent attended and unattended states. All items are recorded in STVRM, and matches between similarly tagged items are facilitated.

Feature-based repetition detection differs dramatically from attentional selection based on spatial or temporal location and from feature-based selection for partial report. Target items can be selected for partial report from stimulus streams based on physical differences in intensity or when targets are surrounded by outline squares.

If feature-based selection for partial report follows STVRM in the same processing path, then the locus of attentional filtering is constrained to lie between STVRM and the process of partial-report selection.

Acknowledgments

The experimental work was supported by Office of Naval Research, Cognitive and Neural Sciences Division, Grant N00014-88-K-0569; the theoretical work and preparation of the manuscript was supported by AFOSR, Life Science Directorate, Visual Information Processing Program; Grant No. 91-0178. For reprints, write: Prof. G. Sperling; HIPLab/Psychol/NYU; 6 Washington Pl., Rm. 980; NY, NY 10003.

References

Averbach, E. & Sperling, G. (1961). Short term storage of information in vision. In C. Cherry (Ed.), *Information Theory*. Washington, D.C.: Butterworths, 1961. Pp. 196-211.

Broadbent, D.E. (1958). *Perception and Communication*. London: Pergamon Press.

Cave, K. R., & Wolfe, J. M. (1990). Modelling the role of parallel processing in visual search *Cognitive Psychology 22*, 225-271.

Deutsch, J.A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review*, **70**, 80-90.

Dosher, B. A., Sperling, G., & Wurst, S. A. (1986). Tradeoffs between stereopsis and proximity luminance covariance. *Vision Research, 26*, 973-990.

Duncan, J., & Humprheys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96* 433-458.

Felfoldy, G. L., & Garner, W. R. (1971). The effects on speeded classification of implicit and explicit instructions regarding redundant dimensions. *Perception and Psychophysics*, **9**, 289-292.

Folk, C.L. & Egeth, H. (1989). Does the identification of simple features require serial processing? *Journal of Experimental Psychology: Human Perception and Performance*, **15**, 97-110.

Garner, W. R. (1978). Selective attention to attributes and to stimuli. *Journal of Experimental Psychology: General*, **107**, 287-308.

Hoffman, J. E. (1979). A two-stage model of visual search. *Perception and Psychophysics*, **25**, 319-327.

Intraub, H. Visual dissociation: An illusory conjunction of pictures and forms. *Journal of Experimental Psychology: Human Perception and Performance, 11*, 431-442.

Kanwisher, N. G., & Potter, M. C. (1989). Repetition blindness: The effects of stimulus modality and spatial displacement. *Memory and Cognition*, **17**, 117-124.

Kanwisher, N. G. (1987). Repetition blindness: Type recognition without token individuation. *Cognition*, **27**, 117-143.

Kaufman, J. *Visual repetition detection*. Unpublished doctoral dissertation, New York University, 1978.

Kinchla, R. A. (1980). The measurement of attention. In R.S. Nickerson (ed.), *Attention and Performance VIII*, 1980, Hillsdale, N.J.: Erlbaum.

Krantz, D. H. & Tversky, A. (1971). Conjoint-measurement analysis of composition rules in psychology. *Psychological Review, 78*, 51-169.

LaBerge, D., & Brown, V. (1989). Theory of attentional operations in shape identification. *Psychological Review, 96*, 01-124.

Landy, M. S., Cohen, Y., & Sperling, G. (1984a). HIPS: Image processing under Unix. Software and applications. *Behavior Research Methods and Instrumentation, 16*, 199-216.

Landy, M. S., Cohen, Y., & Sperling, G. (1984b). HIPS: A Unix-based image processing system. *Computer Vision, Graphics, and Image Processing, 25*, 331-347.

Marohn, K. M., & Hochhaus, L.. (1988). Different case repetition still leads to perceptual blindness. *Bulletin of the Psychonomic Society, 26*, 29-31.

McClelland, J. L., & Rumelhart, D. E. (1988) *Parallel distributed processing. Vols. 1 and 2*. Cambridge, MA: MIT Press/Bradford Books.

Merikle, P. M. (1980). Slection from visual persistence by perceptual groups and category membership. *Journal of Experimental Psychology: General*, **109**, 279-295.

Milner, P. M. (1974). A model for visual shape recognition. *Psychological Review*, **81**, 521-535.

Murdock, B. (1982). A theory for the storage and retrieval of item and associative information.

Psychological Review, **89**, 609-626.
Nakayama, K. & Silverman, G.H. (1986). Serial and parallel processing of visual feature conjunctions. *Nature*, **320**, 264-265.
Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.
Norman, D. A. (1968). Towards a theory of memory and attention. *Psychological Review, 75*, 522-536.
Orenstein, H. B. & Holding, D. H. (1987). Attentional factors in iconic memory and visible persistence. *The Quarterly Journal of Experimental Psychology*, **39A**, 149-166.
Parish, D. H. & Sperling, G. (1991). Object spatial frequencies, retinal spatial frequencies, noise, and the efficiency of letter discrimination. *Vision Research, 31*. 1399-1410.
Pavel, M. (1991). Model of preattentive search *Mathematical Studies in Perception and Cognition, 91-4*, New York University, Department of Psychology, 1991.
Posner, M. I., Poies, S. J., Eichelman, W., & Taylor, Rl L. (1969) Retention of visual and name codes of single letters. *J. Exptl. Psychol. Monograph, 70*, 1-16.
Reeves, A., & Sperling, G. (1986). Attention gating in short-term visual memory. *Psychological Review, 93*, 180-206.
Runtime Library for Psychology Experiments. (1988). N.Y.: HIPLab.
Sagi, D. (1988). The combination of spatial frequency and orientation is effortlessly perceived. *Perception and Psychophysics, 43*, 601-603.
Shiffrin, R. M. & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, **84**, 127-190.
Sperling, G. (1960). The information available in brief visual presentation. *Psychological Monographs*, **74** (11, Whole No. 498).
Sperling, G. (1963). A model for visual memory tasks. *Human Factors, 5*, 19-31.
Sperling, G., Budiansky, J., Spivak, J. G., & Johnson, M. C. (1971). Extremely rapid visual search: the maximum rate of scanning letters for the presence of a numeral. *Science, 174*, 307-311.
Sperling, G., & Dosher, B. A. (1986). Strategy and optimization in human information processing. In K. Boff, L. Kaufman, & J. Thomas (Eds.), *Handbook of Perception and Performance. Vol. 1*. New York, NY: Wiley, 1986. Pp. 2-1 to 2-65.
Sperling, G. & Kaufman, J. (1978). Three kinds of visual short-term memory. Talk presented at Attention and Performance VIII, Educational Testing Service, Princeton New Jersey, August 22, 1978.
Sperling, G. & Kaufman, J. (1991). Visual repetition detection. *Mathematical Studies in Perception and Cognition, 91-1*, New York University, Department of Psychology, 1991.
Sperling, G. & Melchner, M.J. (1978a). Visual search, visual attention, and the attention operating characteristic. In J. Requin (ed.), *Attention and Performance VII*, Hillsdale, N.J.: Erlbaum. Pp. 675-686.
Sperling, G. & Melchner, M. J. (1978b). The attention operating characteristic: Examples from visual search. *Science*, **202**, 315-318.
Sperling, G. & Reeves, A. (1980). Measuring the reaction time of a shift of visual attention. In R. Nickerson (Ed.), *Attention and Performance VIII*. Hillsdale, New Jersey: Erlbaum. Pp. 347-360.
Sperling, G., & Weichselgartner, E. (1991). Episodic theory of the dynamics of spatial attention. *Psychological Review, 91*. (In press.)
Swets, J. (1984). In R. Parasuraman & D. R. Davies (Eds.), *Varieties of Attention*. New York, N. Y.: Academic Press. Pp. 183-242.

Treisman, A. M. (1977). Focused attention in the perception and retrieval of multidimensional stimuli. *Perception and Psychophysics, 22,* 1-11.

Treisman, A. M. (1982). Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 194-214.

Treisman, A. M. (1986). Properties, parts, and objects. In K. R. Boff, L. Kaufman, & J. P. Thomas (eds.), *Handbook of Perception and Human Performance: Vol. II,* New York: Wiley.

Treisman, A. M. & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology,* **12,** 97-136.

Von Wright, J. M. (1968). Selection in visual immediate memory. *Quarterly Journal of Experimental Psychology,* **20,** 62-68.

Weichselgartner, E. (1984). Two processes in visual attention. Unpublished doctoral dissertation. Department of Psychology, New York University, 1984.

Weichselgartner, E., & Sperling, G. (1987). Dynamics of automatic and controlled visual attention. *Science, 238,* 778-780.

Wright, C. E., & Main, A. M. (1991). Selective search for conjunctively-defined visual targets. Manuscript submitted for publication.

Wurst, S. A. (1989). Investigations of short-term visual repetition memory. Unpublished doctoral dissertation, Department of Psychology, New York University.

Wurst, S. A., Sperling, G., & Dosher, B. A. (1991). The locus and process of visual repetition detection. *Mathematical Studies in Perception and Cognition, 91-2,* New York University, Department of Psychology, 1991.

Motion Perception Between Dissimilar Gratings: A Single Channel Theory

Peter Werkhoven,
Charles Chubb *
and
George Sperling

Department of Psychology and Center for Neural Science,
New York University,
New York, New York 10003,
USA.

(PREPRINT)

April 10, 1992

*Present address: Department of Psychology, Rutgers University, New Brunswick, NJ 08903.

Contents

Abstract

We examine apparent motion carried by textural properties. The texture stimuli consist of patches of sinusoidal grating of various spatial frequencies and contrasts. Phases are randomized between frames to insure that standard motion analysis *directly* applied to stimulus *luminance* is not systematically engaged.

We use ambiguous apparent motion displays in which a heterogeneous motion path defined by alternating patches of texture s (standard) and texture v (variable) competes with a homogeneous motion path defined solely by patches of texture s. Our results support a model in which strength of texture-defined motion is computed from a single spatial transformation of the stimulus - the *activity* transformation. The value assigned a point in space-time by the activity transformation is directly proportional to local texture contrast and inversely proportional to local spatial frequency (within the range of spatial frequencies examined). Thus, the activity transformation can be modeled as the rectified output of a low-pass spatial filter applied to stimulus contrast. The strength of texture-defined motion between a patch of texture s and a patch of texture v is proportional to the product of the activities of s and v. A counterintuitive implication of this model borne out in our data is that apparent motion along a heterogeneous path consisting of alternating patches of a low contrast, low frequency texture (texture l) and patches of high contrast, high frequency texture (texture h) can be stronger than motion along a homogeneous path of identical patches of texture h.

1 Introduction

1.1 Terminology

Drifting spatiotemporal modulations of a variety of optical entities (such as luminance, contrast, texture type, binocular disparity, etc.) can induce a vivid motion percept, that is, something appears to move from one place to another. This introspective description, however, does not necessarily reflect the underlying processes in human visual motion perception. We discriminate two stages for the extraction of motion information. A *preprocessing* stage serves to transform the raw stimulus into a single scalar modulation signal (~~making the optical entities explicit~~) and is ~~fed into~~ next stage: *motion analysis* (~~making the motion information explicit~~).

The preprocessing stage can be either a linear or a nonlinear stimulus transformation. *Linear* preprocessing is called *first order* motion extraction, whereas *nonlinear* preprocessing is called *second order* motion extraction (Cavanagh *et al.*, 1989; Chubb and Sperling, 1989).

Both first and second order motion extraction can be further classified by the type of motion analysis. A review of the literature on motion perception shows that two classes of motion extraction mechanisms have been considered and tested experimentally. We call these classes of motion extraction *motion correspondence extraction* and *motion energy extraction*.

Motion *energy* extraction compute the directional energy of a Fourier representation of the drifting modulation signal, that is, the relative energy of 'drifting' spectral components. *Energy extraction* is insensitive to the relative phase of the different spatial Fourier components of the modulation signal (van Santen and Sperling, 1984), and are thus insensitive to features of spatial structure. Motion *correspondence extraction*, in contrast, is thought to identify local features of the modulation signal and

then to track the location of corresponding features over time. Energy and correspondence mechanisms yield qualitatively different predictions for the strength of motion (Werkhoven *et al.*, 1990b).

In this paper, we study drifting modulations of texture *type* (second order motion extraction) and discriminate between motion correspondence extraction and motion energy extraction.

1.2 Motion-From-Texture

Texture-defined motion is motion carried by textural properties. It is not produced by a moving texture patch; that would be rigid, luminance-defined motion. Texture-defined motion is usually produced by a moving patch that is filled with a particular *type* of texture in which each successive frame represents a new, uncorrelated instance of that texture type. As for all second order motion stimuli, an intriguing aspect of texture-defined motion perception is that (unlike perception of luminance defined or first order motion) it cannot be explained by Fourier energy or autocorrelational motion analysis (standard motion analysis). An early example of texture-defined motion was reported by Sperling (1976). Detailed studies and analysis were recently presented by Chubb and Sperling (1989,1991), Cavanagh *et al.* (1989), Lelkens and Koenderink (1984), Mather (1991), Turano and Pantle (1989), and Victor and Conte (1989).

1.3 Energy Channels

Chubb and Sperling (1989, 1991) suggested a scheme for extracting texture-defined motion that consists of two stages. Stage 1 is a 'texture grabber' that consists of a linear spatial filter followed by a non-linearity (*e.g.*, rectification, squaring, *etc.*). Stage 2 is standard motion analysis.

The 'texture grabbers' of stage 1 are assumed to be distributed over the visual field. It is assumed that the spatial filters of stage 1 operate on stimulus contrast (see Model section), rather than on luminance, but this assumption is not critical to our arguments. The output of a linear filter may be positive or negative depending on the local phase of the sensed texture, and that would yield an expected output of zero over the phase-randomized texture patches. The purpose of rectification is to produce a positive average output across the texture so that a texture grabber registers the presence or absence of texture, independent of local phase. Indeed, that is why Stage-1 (linear spatiotemporal filter followed by rectification) is called a *texture grabber*. The output of a texture grabber in response to a particular texture is called *activity*. The essential nonlinear characteristic of texture extraction processes has also been emphasized by Bergen and Adelson (1988) and Caelli (1985) (see also Graham, 1992).

We assume that the second stage (standard motion analysis) is a coincidence detector with an internal delay (*e.g.*, Reichardt, 1961). It computes the following product: previous activity at location 1 times current activity at location 2. Such a motion-detection scheme yields a high response when the time for a texture to move from location 1 to location 2 equals the internal delay. The output of the second stage corresponds to motion strength. This model of standard motion analysis is proposed only for the sake of definiteness. van Santen and Sperling (1984) showed that the various motion models that had been proposed for human vision were equivalent or approximately equivalent to an

elaboration of this simple scheme. We will refer to this class of motion models as *standard motion analysis*. None of our conclusions will depend on the details of standard motion analysis.

Together, a texture grabber followed by standard motion analysis form a likely motion computation, which we call an *energy-channel*. There may exist multiple energy-channels, yielding independent measures of motion strength.

1.4 Correspondence-Channels

Above, we discussed a type of motion computation (energy-channels) that is basically insensitive to *similarities* between the textures in a motion path. Traditionally, however, psychophysicists have interpreted results of motion-from-texture experiments in terms of *correspondence matching*. The metaphore of motion correspondence describes motion as the convection of some invariant aspects of spatiotemporal structure over time. A correspondence computation is inherently different from an activity computation: it is highly sensitive to differences between textures in a motion path. Although it is intuitively clear that a *single* energy-channel is inherently different from a correspondence computation, this may not be intuitively clear for a system with *multiple* energy-channels. However, in the Model section of this paper, we formally show that a system of *multiple* energy-channels can not be equivalent with correspondence matching. We conclude that a correspondence computation is a separate class of motion computations.

Historically, motion correspondence has been investigated with ambiguous motion displays in which motion is perceived as occurring along one or the other of several competing paths. Most studies have dealt with stimuli that stimulated the first-order motion system (*e.g.*, Burt and Sperling, 1981; Kolers, 1972; Navon, 1976; Papathomas *et al.*, 1991; Shechter *et al.*, 1989; Ullman, 1980; Werkhoven *et al.*, 1990a,1990b) and these data are adequately explained by the first-order motion models. We consider here two recent studies that attempt to deal with feature correspondence in texture-defined motion stimuli. These studies illustrate the difficult methodological issues that arise in attempting to determine motion correspondence, and thereby they motivate the more complex paradigm we use.

1.4.1 Watson's crossed-phi procedure

Watson (1986) focussed on the spatial frequency specificity of perception of texture-defined motion. He used a 'crossed phi' method, in which two different texture patches (A, B) exchange position in successive frames (B, A). The pathes were Gaussian-windowed sine waves (Gabor patches). Observers reliably perceived lateral motion when A and B were different spatial frequencies. Watson interpreted his results in terms of models of human visual motion processing in which motion estimates are computed separately within different spatial frequency bands. Furthermore, it was implicitly assumed, that such a model was equivalent to a correspondence computation.

In our view, the ambiguous 'crossed-phi' paradigm admits two alternative interpretations in terms of competing paths. The motion paths between textures of similar spatial frequencies (A - A), (B - B) must compete with a *no-motion* (standstill) path between textures of different spatial frequencies (A,B). Alternatively, the motion path A - A in one direction competes against the motion path B - B

in the opposite direction. Neither interpretation allows the crossed-phi paradigm to measure *motion* strength between textures of different spatial frequencies because such motion is not present in the display.

1.4.2 Green's Gabor patches

Green (1986) embedded competing paths in a rotating annular display (quite similar to Navon, 1976). He measured the strength of apparent motion as a function of the spatial frequency of Gabor patches along the paths. Green's observers tended to perceive motion along paths in which neighboring patches had same or similar spatial frequencies. Green concluded that spatial frequency is a strong determinant in 'correspondence matching' and proposed (like Watson) that the visual system uses *multiple* (bandpass) channels similar to spatial frequency channels in analyzing these texture-defined motion paths.

Green's results are open to a different interpretation. If Green had failed to find dominance of the homogeneous path, it would indeed have indicated a failure of matching in apparent motion; *i.e.*, that all patches were equivalent for motion-from-texture. However, we propose that, in Green's motion competition scheme, finding dominance of the homogeneous motion path does not prove the existence of different channels or of feature-matching. The argument is as follows.

The motion stimulus used by Green contains two homogeneous motion paths (one between patches of texture 1 and another between patches of texture 2) in the same direction and one type of heterogeneous motion path (between patches of textures 1 and 2) in the opposite direction. We apply a simple but general motion computation (after Werkhoven *et al.*, 1990b) to the Navon-Green display to illustrate that motion in the direction of the heterogeneous path should never be perceptually dominant. Consider an arbitrary texture grabber (T) followed by a half-Reichardt motion detector (van Santen and Sperling, 1985). Let T_1 be the response of texture grabber T to texture 1 and T_2 the response of the texture grabber T to texture 2. Using the Reichardt product to estimate motion strength, and subtracting the strength of the heterogeneous from the (opposite) homogeneous path yields a net motion strength in the homogeneous direction proportional to $(T_1 - T_2)^2$ which is always non-negative! The dominance of the homogeneous motion path is inherent to the display's competition scheme rather than a result of 'correspondence matching'.

The Navon-Green ambiguous motion display is useful in determining whether or not two competing texture patches are motion metamers. With metamers, motion is ambiguous; with non-metamers the homogeneous path dominates. However, Green did not elaborate the method sufficiently to demonstrate that more than one type of texture grabber was involved. In particular, Green's results are consistent with the proposal that the visual system uses only a *single* texture grabber in its analysis of texture-defined motion: a texture grabber whose response is monotonic with the spatial frequency of the sensed texture - for example, a texture grabber that applies a *low-pass* spatial filter to stimulus contrast.

1.4.3 Finally

The experiments discussed above on motion-from-texture experiments do not uniquely support a motion correspondence computation. With a different type of experimental paradigm (motion direction discrimination experiments) Victor and Conte (1990) have shown that correspondence might not be relevant to motion processing at all.

To conclude on the validity of correspondence computations, we need a stronger experimental paradigm which is presented in this paper. In the next we will refer to motion correspondence computations as *correspondence-channels.*

1.5 Representation of a General Motion Computation

Above, we have discussed two types of motion computations: energy-channels and correspondence-channels. There may exist an arbitrary number of independent channels of each type contributing to motion strength. In general, the output of the motion computation will be a vector, its components representing the outputs of the individual channels or combinations of them. For example, the motion computation may be represented by a n-dimensional vector when its components correspond to n independent motion channels. However, the motion computation may result in a scalar representation, when all channels are combined (for example summed) before the final representation. The dimensionality of the vector representation is called the dimensionality of the motion computation.

It is important to realize that the dimensionality of the motion computation is the number of values through which all channels involved are finally represented. It does *not* indicate the *number* of channels involved. For example, n independent energy-channels may be represented by a *single* scalar: the sum of all channels.

1.6 Fundamental Questions

In this paper we address the following questions:

- (a) What is the dimensionality of the motion computation? That is, if the motion computation is represented by a vector, what is the dimensionality of this vector.
- (b) What is the number of channels and how are they combined in the final motion vector representation? For example, how many energy-channels (or texture grabbers) are involved in the perception of motion-from-texture?
- (c) What are the characteristics of the channels involved? For example, what are the textural properties sensed by the texture grabbers and how does the activity of a texture grabber depend on these textural properties?

1.7 Motion Metamers

The concept of motion metamers is critical to our method of answering the questions concerning the dimensionality of motion perception. Motion metamers are understood most easily by analogy

with color metamers. Color metamers are patches of light that are judged equivalent with respect to color but which may have different spectral compositions. The explanation for color metamers derives from the standard Young-Helmholtz trichromatic color theory. For human vision, any color can be represented by three independent scalars (Helmhotz, 1924). Equivalently, any arbitrary color can be matched by a suitably chosen mixture of three independent primary colors. In fact, *any* triplet of independent primary colors can serve as a set of primaries for matching an arbitrarily chosen color. The class of all triplets that match a particular color is the metameric class for that color. In order to find such a metameric class of matches for an n-dimensional color system, the number of 'primary colors' used to match an arbitrary 'color' must be at least equal to n.

Richards (1979) applied the matching methods of colorimetry to study a wide range of sensory attributes such as visual flicker, visual texture, tactile sensations, and auditory sensations. Williams *et al.* (1991) observed a kind motion metamerism: polymorphic motion stimuli that produce motion percepts with equivalent directional characteristics, but they did not interpret their observations in terms of the metamerism. Here, we apply the dimensional analysis of colorimetry to the motion domain, in particular, to motion-from-texture.

We say that two kinds of texture patches are motion metamers if they can be interchanged in any motion path without affecting the the motion strength of path. Our experiments are basically analogous to the color matching experiments. First, we embed a variable texture v in a motion path that is in competition with a standard path containing a texture s. We determine the parameters of v (such as contrast and spatial frequency) that are needed to achieve path equality for v and s. Second, we determine that v and s are indeed equivalent for other paths as well. Generalized path equivalence demonstrates that v and s are motion metameres. Third, we find the range of v texture patches that are equivalent to s, *i.e.*, we find the metameric class containing s. From the dimesionality of the metameric class, we can determine the number of 'primary textures' *vsubi* that would be needed to provide a match to every member of the class. That is, different texture primaries might have to be added to provide motion equivalence, just as different color primaries have to be added to provide color equivalence. This number of required texture primaries is the dimensionality of the motion-from-texture strength computation. Fourth, it is generally desirable to repeat the procedure for different choices of s to assure that s was not an unlucky choice. Given our particular results, this iteration of steps 1-3 will not be crucial.

In general, the number of types of texture grabbers is at least as great as the dimensionality of the motion-from-texture computation, and it may be much larger. We offer a proof that, under certain assumptions, the dimensionality of the motion-from-texture computation exactly describes the number of types of texture grabbers involved in the motion-from-texture computation (just as the dimensionality of color space describes the number of types of color receptors).

Unlike color metamers, however, which are equivalent in all subsequent processing, motion metamers are equivalent only with respect to the motion-from-texture computation, and generally are perceived as different in other respects.

1.8 Motion Competition Schemes

The matching technique could basically be applied to a variety of ambiguous motion schemes for determining the *dimensionality* of the motion computation. However, not all of them have the power to discriminate between different type of motion channels (see for example the discussion on Green's display). We used an ambiguous motion scheme that was first introduced by Werkhoven *et al.* (1990). In this motion competition scheme, *one* heterogeneous motion path (between patches of texture s and texture v) competes *directly* with *one* homogeneous path (between patches of texture s).

By varying the textural properties of the textures v, we can determine the heterogeneous motion paths $s - v$ that are metamer with a certain homogeneous path $s - s$. In general, this metamery does not transfer to the textures itself. That is, for a general motion computation, a metamery between motion *paths* $s-s$ and $s-v$ does *not* imply a metamery between *textures* s and v with respect to motion processing. However, it should be noted that, if and only if motion-from-texture is ruled by a *single* energy-channel a metamery between motion paths *does* imply a metamery between the textures with respect to motion processing. This is inherent to energy-channels: textures contribute independently to motion strength.

This competition scheme not only allows to determine the dimensionality of the motion computation, but also allows to determine the number and type (activity versus correspondence) of channels involved in the motion computation. This requires a thorough analysis (given in the Model section).

However, an intuitively clear property of this scheme is that the two type of motion channels considered above (activity versus correspondence-channels) yield qualitatively different predictions for motion metamery and the relative strength of the heterogeneous and homogeneous motion paths. Hence, they are easily discriminated.

1.8.1 Activity-Channels

Consider, for example, a single energy-channel that uses only a single type of texture grabber (*e.g.*, a low pass filter followed by rectification). This model yields the counterintuitive prediction that the motion strength between a patch of high spatial frequency texture and a patch of low spatial frequency texture (heterogeneous motion) can be stronger than motion between two patches of high spatial frequency texture (homogeneous motion). Indeed, such a heterogeneous motion path may dominate homogeneous paths even for a system with multiple energy-channels (when there is more than one type of texture grabber).

For more generality, we apply the energy-channel computation (described above to show the inherent dominance of the homogeneous paths in Greens' display's) to the competition scheme used in this paper. The motion strength is T_1T_2 for the heterogeneous path and T_2T_2 for the homogeneous path yielding a net motion strength in the heterogeneous direction equal to $T_2(T_1 - T_2)$. Whenever T_1 is larger than T_2, the heterogeneous motion path dominates. The motion paths are balanced when T_1 equals T_2. All textures that transform into the same activity are metameric with respect to motion processing!

1.8.2 Correspondence-Channels

By definition, correspondence-channels favor the homogeneous motion path (between patches of similar texture). Consequently, heterogeneous motion cannot dominate over the homogeneous motion path!

1.9 A Preview

1.9.1 Dimensionality of the Computation

In this paper, we discuss a general motion computation consisting on n energy-channels, and a correspondence-channel. By studying the above competition scheme with many different pairs of texture patches (Experiment 1 and 2), we can determine classes of motion metamers and infer the dimensionality of the motion computation (Model section). The results strongly support a one-dimensional motion computation. That is, motion strength is represented by a *single* scalar.

1.9.2 Type of Computation

The experimental results show *no* involvement of a correspondence-channel. Furthermore, the competition schemes allow to determine the *number* of energy-channels that are represented by a single scalar. We proof (under certain assumptions) that the motion computation consists of a *single* energy-channel, that is, a *single* texture grabber followed by standard motion analysis.

Also, we derive the characteristics of the single texture grabber involved with respect to textural properties relevant to motion strength (using the results of Experiment 1, 2 and 3). Thus we can predict the strength of texture-defined motion as a function of the contrast and spatial frequency of sinusoidal texture patches.

1.9.3 Where in the Visual System?

The results of Experiments 4 and 5 will shed some light on where in the stream of visual processing our proposed texture-from-motion computation takes place. In Experiment 4 we show that the motion computation is not based on perceived contrast. Experiment 5 shows that our texture-from-motion stimuli give similar results for monocular, binocular and dichoptic presentations.

2 Method

In this section we describe the ambiguous motion competition scheme used in the experiments.

This scheme (proposed by Werkhoven *et al.*, 1990 b) differs from other schemes (*e.g.*, Burt and Sperling, 1981; Green, 1986; Navon, 1976; Shechter *et al.*, 1989; Ullman, 1980) in that it contains a *single* heterogeneous motion path (between patches of texture 1 and texture 2) that competes *directly* with a *single* homogeneous motion path (between identical patches of texture 2). Except for textural properties, the other parameters (such as step size and frame rate) of the motion paths are identical.

Instead of varying both textures 1 and 2, we sampled a subspace of possible textures resulting in two (similar) schemes: Scheme I and Scheme II. In Scheme I, we kept texture 2 constant (called

texture s) and varied texture 1 (called texture v). In Scheme II, we kept texture 1 constant (now texture s) and varied texture 2 (now texture v).

2.1 Stimulus

2.1.1 Motion Competition Scheme I

In Experiment 1, we used motion competition Scheme I. The motion stimulus consisted of a series of 8 frames ($f_1, f_2, \ldots f_8$) shown successively in time. Figure 1 shows a sketch of the frames.

— Figure 1 about here —

The first frame (f_1) contains an annulus of patches of alternating texture types s and v at regular positions (see Fig. 1, at the left side). Because the viewing distance was constant throughout the experiment, we will specify dimensions in degrees of visual angle. The annulus of texture patches has an inner radius of $r_1 = 1.04$ deg, and an outer radius of $r_2 = 2.08$ deg. The mean radius r is 1.56 deg. The patches (or *sectors*) are spatially contiguous. Since the annulus contains 8 sectors, each sector has a width of 45 deg.

Frame f_2 was similar to frame f_1, except that patches of texture v are replaced by a uniform patch of background luminance. Furthermore, f_2 was rotated around the center of the annulus 22.5 degrees with respect to frame 1 (see Fig. 1, left).

In a sequence of frames, the locations and types of patches in frame f_{n+2} were identical to frame f_n , except for a rotation around fixation of 45 deg.

The presentation time of a single frame ('frame-time') was 133.3 msec. Thus, the presentation time of the 8-frame sequence was 1.066 sec. The annulus revolved at an angular speed of 168.8 degrees per second, yielding a local velocity of the patch-centers of 4.6 degrees of visual angle per second.

The ambiguous motion stimulus described above contains two motion paths. This can be understood most easily using a diagram in which we show the angular positions (φ) of the patches of texture for successive frames. Angular position is measured clockwise relative to the vertical. Such a diagram is shown in Fig. 1, at the right side. Note that the horizontal rows of patches correspond to frame 1,2,3 and 4 respectively. By definition, motion extraction is based on the dynamic properties of the stimulus, that is the spatiotemporal pattern of textures. In the diagram, possible motion paths are spatiotemporal (oblique) rows of elements. The arrows pointing to the left and right are examples of motion paths to the left and right respectively. In the following description of the stimulus, we will say that the neighboring elements in a motion path are spatiotemporally linked or 'matched'. Note that the term 'matching' is used for the purpose of stimulus description only and that it does *not* refer to a 'motion correspondence' computation.

When frame f_n and frame f_{n+1} were presented in succession, two matches between patches of frame f_n and patches of frame f_{n+1} were *a priori* possible. The first match is a homogeneous clockwise match between patches of identical texture s separated by +22.5 deg (indicated in the diagram by the

arrow pointing down and to the right). The second match is a heterogeneous counter-clockwise match between patches of texture v and patches of texture s (-22.5 deg, indicated by the arrow pointing down and to the left). Matches between frames f_n and f_{n+2} are entirely ambiguous. Matches between patches of frames f_n and f_{n+3} involve large temporal separations (400 msec) relative to the equivalent matches between frames f_n and f_{n+1} (133.3 msec). It has been shown that motion strength decreases strongly and monotonically with temporal interval for intervals larger than approximately 30 ms (Burt and Sperling, 1981; Werkhoven *et al.*, 1991). Therefore, the matches between frames f_n and f_{n+3} are unimportant for motion perception in these stimuli.

Scheme I displays contain homogeneous and heterogeneous motion paths in opposite directions. By randomizing the direction of rotation, the directions of the two motion paths (although still opposite) are randomized.

The annular pinwheel stimulus was used for various reasons. First, the motion stimulus was presented at a constant eccentricity in the parafovea, and the effects of anisotropy of the retina were averaged across equivalent areas of the visual field. Second, it was easier to maintain fixation so eye movements were better controlled[1]. Finally (with the use of circularly symmetric stimuli) a motion path does not end at the boundaries of the display, avoiding edge effects.

2.1.2 Motion Competition Scheme II

Scheme II (used in Experiment 2) is equivalent to Scheme I, except that textures s and v are interchanged. The motion stimulus and resulting motion paths for this experiment are sketched in Fig. 2.

-- Figure 2 about here --

Although the heterogeneous motion path (between patches of texture s and v is identical to that of Scheme I, the homogeneous motion path is different from that of Scheme I. In Scheme II, the homogeneous motion path consists of patches of texture v. The critical importance of the two schemes for our paradigm concerns the question of whether, when a particular s and v are chosen so that motion paths are balanced in Scheme I, the paths will remain balanced when the same s and v are used in Scheme II. From the subjects' point of view, however, there is no difference between the two schemes because, for any stimulus generated by Scheme I, an identical stimulus can be generated by Scheme II. However, during the course of a session, when v is varied between trials, different families of stimuli are generated by the two schemes.

[1]Torsional eye-movements induced by the rotating annuli (cyclo-induction) were not controlled in our experiment. Balliet and Nakayama (1978) reported the ability of extremely trained subjects to make stepwise eye torsions up to rotations of approximately 26 degrees for large field stimuli (25-50 degrees of visual angle). However, we do not expect torsional pursuit in our experimental conditions: small field stimuli, brief presentations, fast motion, unpredictable motion direction, and ambiguous or near-threshold motion stimuli.

2.2 Texture Stimuli

The textures used to characterize texture-defined motion are patches of sinusoidally modulated gratings that differ in spatial frequency and contrast. The grating patches were arranged in eight sectors of an annulus (pinwheel) around the fixation point with the grating extending radially in each sector. Two critical parameters that characterize a texture patch at a given location of the pinwheel are contrast c and spatial frequency ω. Within a location, grating orientation was always radial. The phase γ of the grating was a random variable with a uniform distribution.

We use polar coordinates to further characterize the pinwheel. Let φ be the polar angle of a point in the image, and ρ be the distance to the origin (the center of the annulus). Then the luminance distribution at the point ρ, φ in sector j of frame i is:

$$L_{i,j}(\rho,\varphi) = L_0[1 + c_{i,j}\sin(2\pi r\varphi\omega_{i,j} + \gamma_{i,j})]. \quad (1)$$

We define the mean spatial frequency $\omega_{i,j}$ as the spatial frequency at mean radius r. The mean spatial frequency $\omega_{i,j}$ of a texture patch depends only on whether j is odd or even. That is, two spatial frequencies, ω_s, ω_v strictly alternate between adjacent patches on every frame of the display.

Within a trial, the contrast $c_{i,j}$ of a sector i,j depended only on whether i and j were even or odd. On odd frames, $c_{o,j}$ was chosen as c_s or c_v according to whether the sector j was even or odd. On even frames, sector contrast $c_{e,j}$ alternated between 0 and c_s in Scheme I and between c_v and 0 in Scheme II. Between trials, c_v and ω_v were changed. Sixteen values of contrast c_v from 0 to 1 were used increasing by steps of 0.0625: 0, 0.0625, 0.13 ..., 1. Spatial frequency ω_v was varied over a range of three octaves: 1.2, 2.5, 3.7, 4.3, 4.9, 5.6, 7.4 or 9.9 cpd. The contrast c_s and spatial frequency ω_s of texture s were constant throughout the experiment: $c_s = 0.5$, $\omega_s = 4.9$ cpd.

The phase $\gamma_{i,j}$, $0 \le \gamma_{i,j} \le 2\pi$, was chosen randomly and independently for every combination of i and j, that is, for every single patch. The phase randomization of every patch makes the motion of the stimulus inaccessible to any first-order (Fourier-based) mechanism. Phase randomization insures that motion mechanisms sensitive to correspondences in stimulus luminance were not systematically engaged (Chubb and Sperling, 1988).

— *Figure 3 about here* —

Figure 3 shows an example of a series of frames for Scheme I. Texture s is a 'medium' frequency grating and texture v is a 'low' frequency grating. The regions inside and outside the annulus (background) were uniform grey and had a luminance value ($L_0 = 72$ cd/m^2). Within the annulus' texture patches the expected luminance value was equal to the background luminance.

2.3 Apparatus

The experiment was controlled by a IBM 386 PC compatible computer, driving a TrueVision AT-Vista video graphics adapter. A 60 Hz Imtec 1261L monitor with a P4-type phosphor was used to

display the stimuli. The screen dimensions were 21.8 x 14 cm (640 x 480 pixels; 12.3 x 8.0 deg visual angle)[2]. We used a look-up table to linearize the monitor's luminance values with the gray values of the computed stimulus patterns. The decay time to 10% and 1% intensity was about 1.3 and 6.2 msec respectively which is shorter than the temporal properties of retinal processing (Farrell *et al.*, 1990; Sperling, 1971).

2.4 Subjects

Two subjects participated in the experiments: one of the authors (PW) and a colleague (JS). PW is emmetropic. JS is myopic (-0.5 D) but was in focus for the viewing distance used. Both subjects were experienced psychophysical observers. Natural pupils, binocular viewing, and spectacle corrections were used throughout. Several naive subjects confirmed the main findings for the experiments.

2.5 Procedure

Subjects indicated the dominant motion path (counter-clockwise/clockwise) by pressing one of two buttons. In both experiments, texture s (the *standard* texture) had contrast $c_s = 0.5$ and spatial frequency $\omega_s = 4.9$ cpd. From trial-to-trial, the spatial frequency ω_v and contrast c_v of texture v was varied. The experiments determined the probability $P_i(c_v;\omega_v)$ of perceptual dominance of the heterogeneous motion path as a function of c_v for certain ω_v using the method of constant stimuli. The subscript i, i=1,2, indicates Experiment 1 with competition Scheme I (Fig. 1) or Experiment 2 with Scheme II (Fig. 2).

The probabilities $P_1(c_v;\omega_v)$ and $P_2(c_v\omega_v)$ are estimated by the fraction of perceptually dominant heterogeneous motion paths out of 36 presentations. Spatial frequency ω_v was varied over a range of three octaves: ω_v = 1.2, 2.5, 3.7, 4.3, 4.9, 5.6, 7.4 and 9.9 cpd. Within a session, contrast c_s was varied (pseudo-randomly from trial-to-trial; ω_v was varied only between sessions. For each spatial frequency ω_v. Experiments 1 and 2 were both conducted within one session.

Subjects viewed the stimuli in a room with dimmed background illumination.

3 Experiment 1: Scheme I

3.1 Results

By definition, the homogeneous path (consisting entirely of identical patches of texture s) does not change in this experiment when texture v is varied (see Scheme I, Fig. 1). The strength of the heterogeneous path, which is composed of alternate patches of textures s and v) is varied by varying spatial frequency and contrast, ω_v and c_v, of texture v. Figure 4 shows the probability $P_1(c_v;\omega_v)$ of

[2]Due to the limited bandwidth of the video amplifier (30 MHz) of the monitor, an anisotropy was observed for the average luminance of differently oriented textures that contain high spatial frequencies. Therefore, we only displayed the pixels at column position m and row position n for which $(m+n)$ was even. The other pixels were dark. Hence, vertical and horizontal gratings share a common 'carrier' component. This procedure forfeits maximum luminance and resolution in favor of eliminating anisotropy; the net resolution (320 x 240 pixels) was more than adequate for the displays.

reporting the heterogeneous motion path as dominant as a function of the contrast c_v of texture v. Each panel shows $P_1(c_v;\omega_v)$ for a different value of spatial frequency ω_v.

— *Figure 4 about here* —

The data show that the probability of reporting the heterogeneous path as dominant increases monotonically from zero (for small c_v) to one (for $c_v = 1$) for all values of ω_v except the highest, where the probability of heterogeneous motion dominance has only reached about 65% when $c_v = 1$. A remarkable feature of these data is that in all eight panels, the probability $P_1(c_v;\omega_v)$ of heterogeneous motion dominance exceeds 50% for sufficiently high contrast of c_v.

The upper left panel of Fig. 4 shows data for a two octave difference between the spatial frequency of texture s ($\omega_s = 4.9$ cpd) and the spatial frequency of texture v ($\omega_v = 1.2$ cpd). Heterogeneous motion is perceived in 50% of the presentations when the contrast c_v of texture v is approximately 0.2. Note that at this balance point where both paths are equally likely, both the contrasts and the spatial frequencies of textures s and v are markedly different. Once c_v exceeds 0.5, the heterogeneous motion path is dominant in 100% of the presentations. A 100% perceptual dominance of a heterogeneous over a homogeneous path demonstrates that the similarity between the textures in a motion path certainly is not essential for motion strength. Indeed, for sufficiently large c_v, the heterogeneous path is dominant over the homogeneous path for every combination of frequencies tested in Fig. 4.

The transition contrasts between heterogeneous and homogeneous motion occur where the curves of Fig. 4 cross 50%. The transition contrasts occur at a wide range of different contrasts c_v for different spatial frequencies ω_v. Each P_1 curve is well characterized by two parameters: the *transition contrast* $\mu_1(\omega_v)$ and the *steepness* $\sigma_1(\omega_v)$ at the transition contrast (the subscript 1 indicates Scheme I). The transition contrast $\mu_1(\omega_v)$ is defined as the contrast c_v of texture v, necessary for balancing the motion paths (such that $P_1(c_v;\omega_v) = 50\%$), The steepness $\sigma_1(\omega_v)$ is defined as the derivative $\frac{\partial}{\partial c_v} P_1(c_v;\omega_v)$ with respect to c_v at the transition contrast.

To estimate transition contrast $\mu_1(\omega_v)$ and steepness $\sigma_1(\omega_v)$, we selected[3] data points of each probability curve around the transition contrast. Within this selected range, the curve was assumed to be linear, and these data points were subject to a least square method of linear regression to estimate the regression coefficients $\mu_1(\omega_v)$ and $\sigma_1(\omega_v)$.

— *Figure 5 about here* —

[3]In principle, we selected the three data-points around the transition contrast (the crossing of the curves with the 50% guide line) that were closest to the 50% guide line. There were only two exceptions. First, at spatial frequency $\omega_v = 1.2$ cpd, for subject PW, Experiment 2, we selected the data points with contrast $c_v = 0.19, 0.25$ and 0.31 (to avoid the low contrast values, for which Scheme II becomes ambiguous). Second, at spatial frequency $\omega_v = 2.5$ cpd, for subject JS, experiment I and II, we selected the data points with contrast $c_v = 0.38$ and 0.5 (since we had no data points close to the guide line).

Estimates of $\mu_1(\omega_v)$ are shown in Fig. 5 as a function of the varied spatial frequency ω_v (open circles). The transition contrast $\mu_1(\omega_v)$ increases systematically with increasing spatial frequency ω_v of texture v for both subjects. Together, the data of Figs. 4 and 5 indicate that the strength of the heterogeneous motion path increases with increasing contrast c_v but decreases with increasing spatial frequency ω_v.

— Figure 6 about here —

Estimates of $\sigma_1(\omega_v)$ are shown in Fig. 6 as a function of the varied spatial frequency ω_v (open circles). The *steepness* $\sigma_1(\omega_v)$ of the probability curves at transition contrast $\mu_1(\omega_v)$ decreases with the spatial frequency ω_v of texture v. In the Model section we elaborates on this finding.

3.2 Discussion

3.2.1 A One-Dimensional Motion Computation

We found a metameric class of heterogeneous motion paths $s - v$ that have a strength equal to the homogeneous path $s - s$. For all patches v examined, we only had to adjust the contrast of v to make path $s - v$ match path $s - s$.

Consider the analogy of color matches in scotopic (dark-adapted) vision with metameric motion matches. In scotopic vision, a patch of a standard wavelength (say, 500 nm, or any other visible wavelength) can be matched to a patch of any other wavelength by adjusting the intensity of the standard. Indeed, in scotopic vision, a patch composed of any combination of wavelengths can be matched by a suitably chosen intensity of the standard. Intensity of the standard patch is obviously a one-dimensional continuum. Because, any patch of any visible wavelength can be mapped into an equivalent standard (*e.g.*, 500 nm) intensity, scotopic 'color' vision is one-dimensional.

In our motion-from-texture experiments, the motion path of standard textures s plays the same role as the standard wavelength in scotopic color vision. We obtained motion equivalence between the path composed of standard textures s (frequency $\omega_s = 4.9$ cpd) and paths composed of textures s and v (frequencies ω_v ranging from 1.2 to 9.9 cpd). There was one difference, however, that does not change the logic of the procedure but greatly improves it in practice. To obtain motion equivalence, we chose to vary the contrast of path $s - v$ rather than of the standard path $s - s$. Whereas varying the contrast of the texture (v) is not conceptually different than varying the contrast of the standard s, it has the great experimental advantage of yielding a large class of stimuli ($s - v$ paths with $\omega_v = 1.2 \ldots 9.9$ cpd) all of which are metamers. The stimuli span a two-dimensional space: spatial frequency and contrast. The motion-from-texture computation is one-dimensional because the subject can obtain motion equivalence between any pair of stimuli by turning only one dial–the contrast of one of the stimuli (the contrast of v).

The observed class of metamer motion paths implies a *one-dimensional* motion computation. That is, whatever motion channels are involved in the motion computation, their collective result is represented by a single scalar! However, the required contrast values for the different textures v to

balance the motion paths $s-s$ and $s-v$, can still be determined by multiple channels, including other energy-channels (multiple texture grabbers) or a secondary contribution of a correspondence-channel.

First, we will show that a *single* energy-channel is *sufficient* to model the results, found for Experiment 1. Second, we will discuss how the effects of a possible correspondence-channel can be isolated, using Scheme II.

3.2.2 Sufficiency of A Single Activity-Channel

In a single energy-channel, we assume that only one single type of texture grabber operates on the input yielding an *activity representation* of the input. Motion strength is the result of a standard motion analysis scheme applied to this activity representation. The motion strength of a path is computed from the product of activity measures between successive patches along the path in space-time. Motion strength of a heterogeneous path balances homogeneous motion strength when the responses (activities) to textures v and s are equal. Differences in textural properties between elements s and v are irrelevant as long as the activities are equal, just as, in scotopic vision, differences in wavelength are irrelevant as long as the rod response is the same.

The results for Scheme I suggest an activity transformation that is a monotonically increasing function of contrast and a monotonically decreasing function of spatial frequency. For example, to balance the activity of texture s, with contrast c_s and spatial frequency ω_s, with a lower spatial frequency texture v, (c_v, ω_v) requires a $c_v < c_s$. This pattern of results suggests a single class of texture grabbers consisting of a low-pass spatial filter followed by rectification.

We argued that a single energy-channel is *sufficient* to explain the results of Experiment 1. It is important to note here, however, that our finding that heterogeneous motion can dominate homogeneous motion is also consistent with multi energy-channels, as will be shown in the Model section. For example, the dominance of heterogeneous motion may well be the result of two independent energy-channels, both favoring heterogeneous motion. To uniquely determine the number of channels involved, we need the results for competiton scheme II together with a formal analysis (Model section).

3.2.3 Secondary Contributions of a Correspondence-Channel

In the discussion above, we argued that a *single*-channel model is sufficient to model the (contrast/frequency dependent) dominance of heterogeneous motion found for Scheme I. However, we can not exclude a possible *secondary* effect of texture similarity based on this scheme. To motivate Experiment 2, we need to elaborate on this argument.

Although motion perception may be dominated by a single energy-channel, there may yet be a secondary contribution of a correspondence-channel.

The relative strength of the heterogeneous motion path would decrease as the differences between the spatial frequencies and contrasts of successive patches of textures s and v increased. Suppose there were a secondary contribution of a correspondence-channel. in Experiment 1, sensitive to differences between textures in either contrast or frequency. Because the correspondence-channel favors the homogeneous path (by definition), motion balance requires v in the heterogeneous path to have a

higher contrast c_v to overcome the similarity in path $s - s$ than if there were no correspondence-channel. Thus, in Scheme I, a secondary correspondence effect would displace transition contrast $\mu_1(\omega_v)$ to higher values.

To test for a correspondence-channel, we introduce Scheme II in which s and v are interchanged (see Fig. 2). If there were a correspondence effect, in Scheme II it would favor the v-v path and the transition contrast $\mu_2(\omega_v)$ would be shifted below $\mu_1(\omega_v)$ for any texture v.

When the homogeneous and heterogeneous motion paths remain balanced after interchanging textures s and v, this is called *transition invariance*. Transition invariance would imply that there is *no* contribution of a correspondence-channel.

4 Experiment 2: Scheme II

4.1 Results

Figure 4 shows the probabilities $P_2(c_v;\omega_v)$ of the dominance of the heterogeneous motion path as a function of the contrast c_v of texture v for different spatial frequencies ω_v of texture v. The data points for Scheme II are marked by a filled circle.

When c_v equals zero, the display is physically as well as perceptually ambiguous. A value of 50% is shown for $c_v = 0$, though no data were collected at this point. By varying the contrast of texture v in this experiment, the strength of both the heterogeneous motion path and the homogeneous motion path are varied. As the contrast c_v increases, the probability of heterogeneous motion dominance first increases to a maximum, then decreases to zero for high contrast c_v. On the whole, for contrasts above 0.1 or, in a few cases, 0.2, the Scheme I and Scheme II curves are mirror complementary, and seem to cross at exactly P= 50%. That is, the two schemes produce remarkably similar transition contrasts.

To examine the correspondence between the data from Schemes I and II, some definitions are needed. Let the transition contrast $\mu_2(\omega_v)$ be the contrast c_v of texture v for which the motion paths are balanced, and the probability of heterogeneous motion dominance $P_2(c_v;\omega_v)$ is 50% (the index 2 indicates Scheme II). The steepness at this transition contrast is $\sigma_2(\omega_v)$. The transition contrast $\mu_2(\omega_v)$ and steepness value $\sigma_2(\omega_v)$ are estimated as $\mu_1(\omega_v)$ and $\sigma_1(\omega_v)$ in the previous section.

To compare the transition contrast $\mu_2(\omega_v)$ for Scheme II with transition contrast $\mu_1(\omega_v)$ for Scheme I, they are presented together as a function of spatial frequency ω_v in Fig. 5. Transitions $\mu_2(\omega_v)$ are presented with filled circles. As in Scheme I, the contrast $\mu_2(\omega_v)$ of texture v, necessary for balancing the motion paths, increases systematically with increasing spatial frequency ω_v of texture v. An exception for both subjects are the transition contrasts for ω_v = 9.9 cpd.

To compare the steepness values $\sigma_2(\omega_v)$ for Scheme II with steepness values $\sigma_1(\omega_v)$ (for Scheme I), the absolute value of $\sigma_2(\omega_v)$ is shown as a function of the varied spatial frequency ω_v in Fig. 6 (using filled circles). It should be noted that the estimation is not very accurate: the standard deviation in the distribution of steepness coefficient $\sigma_i(\omega_v)$ is approximately 20%. However, like $\sigma_1(\omega_v)$, the steepness $\sigma_2(\omega_v)$ shows a tendency to decrease with increasing spatial frequency ω_v of texture v.

4.2 Discussion

4.2.1 Transition Invariance and Motion Metamers

It is immediately clear that, for most spatial frequencies ω_v of texture v, the transition contrast $\mu_2(\omega_v)$ is equal within measurement error to transition contrast $\mu_1(\omega_v)$ (see Fig. 5). In 14 of 16 cases, the transition contrasts are invariant when the textures s and v are interchanged. This we call *transition invariance.*

In two cases (the highest spatial frequency used – ω_v = 9.9 cpd– for both subjects), a small difference between transition contrasts for Scheme I and II is observed. At the high spatial frequency of v, the contrast of texture v necessary to balance the motion paths is slightly smaller for Scheme II than for Scheme I. This shift in transition contrast suggests a small similarity effect (a small contribution of a correspondence-channel), and was discussed in the discussion of Experiment 1.

Transition invariance implies that textures s and v (at transitions) are equivalent with respect to motion precessing and can be interchanged in any motion path (Scheme I and Scheme II) without affecting motion strength. This leads to the important conclusion that the metamery for *motion paths* $s-s$, $s-v$ and $v-v$ transfers to the metamery of *textures* with respect to motion processing.

It is interesting to note that Green (1986, Fig. 7, p. 604) was unable to find a contrast that could make a spatial frequency patch of 5.0 cpd into a motion metamer of a 1.7 cpd patch. We had no difficulty in finding metamers between even more disparate spatial frequencies. However, our data in Fig. 5 show that one of the two subjects would require the 5 cpd stimulus to have more than 2x the contrast of the 1.7 cpd stimulus, and this is outside the range of contrasts that Green explored.

4.2.2 Necessity of a Single Acitivity-Channel

The general finding of motion metamery and transition invariance strongly constraints the possible type of motion computations. First, the finding of a class of metameric motion paths indicates that the motion computation is one-dimensional (see Discussion Experiment 1). That is, the motion channels possibly involved are combined and represented by a *single* scalar.

Second, transition invariance shows there is *no* secondary contribution of correspondence-channels (see the discussion on this issue in Experiment 1). The effect that a patch of texture v has on the strength of motion is independent of the other patches in the path. At a transition, the strength of motion path $s-v$ is equal to that of $v-v$ and that of $s-s$, although a correspondence-channel would yield stronger motion for the homogeneous paths.

The two constraints above leave us with a system of multiple energy-channels that must be combined and represented by a single scalar representation (*e.g.*, summation of energy-channels). In the Model Section, we prove (under the assumption of channel summation) that if multiple energy-channels were involved, the transition contrast would generally shift when the textures s and v are interchanged in Schemes I and II. However, when motion perception is exclusively ruled by a *single* energy-channel (the product of the activity of a single type of texture grabber), the transition contrast is invariant when the textures s and v are interchanged. Hence, transition invariance uniquely supports a *single* energy-channel.

5 Experiment 3: Contrast Linearity

5.1 Motivation

In the above experiments, we have shown that the transition contrast $\mu_1(\omega_v)$ increases systematically with increasing spatial frequency ω_v of texture v for both subjects. The strength of the heterogeneous motion path in Scheme I increases monotonically with increasing contrast c_v but decreases with increasing spatial frequency ω_v. In order to further specify the dependency of motion strength on contrast, we performed an experiment similar to that described above using competition Scheme I, and varied the contrast of texture s.

5.2 Results

We kept the frequency of textures s and v constant (ω_s = 4.8 cpd and ω_v = 1.2 cpd) and measured the transition contrast μ_1 as a function of contrast c_s (Scheme I). Transition contrast was estimated from the psychometric curves using the method described earlier.

— *Figure 7 about here* —

Figure 7 shows the transition contrast μ of texture v for three contrast values of texture s (c_s = 0.50, 0.75 and 1.00) for three subjects. The data strongly suggest a *linear* dependence of the transition contrast of texture v on the contrast of texture s. The solid lines are the best fits (minimizing the sum of squares), accounting for at least 97% of the variance for each subject.

5.3 Discussion

We showed that the transition contrast of texture v needed to balance the motion path $s - v$ with the motion path $s - s$ varied linearly with the contrast of texture s. This dependency is easily accommodated in a model where the texture grabber is linear in the contrast of the texture. In fact, one can easily show that contrast linearity follows directly from the linear data under the assumption that the texture grabber is a separable function of spatial frequency and contrast. A linear (low-pass) spatial frequency filter is a simple example of such a separable filter characteristic.

6 Model

6.1 Summary of Model Constraints

We used the analogy with colorimetry and some general assumptions about the possible motion computations involved to reach the conclusion that texture-from-motion strength is ruled by a single energy-channel. We summarize our reasoning.

We discriminate two classes of motion computations: energy-channels and correspondence-channels, yielding different metrics for the strength of a motion path. Consider, a heterogeneous motion path composed of patches of texture s and v. The strength of an *energy-channel* for an $s - v$ path is determined by the product of the activity of texture s and that of v. The activity of a texture is the output of some nonlinear transformation (texture grabber) that maps texture into a scalar. Activity-channels are insensitive to differences in textural properties and allow heterogeneous motion paths $s - v$ to dominate over homogeneous paths. By definition, the strength of a *correspondence-channel* is determined by the similarity of the textural properties of textures s and v. That is, homogeneous paths $s - s$ and $v - v$ dominate heterogeneous paths $s - v$.

In theory, multiple channels of each type may be involved in a motion computation yielding a motion strength vector representation of arbitrary dimensionality. However, the experimental results impose the following constraints. First, the class of metameric motion paths in both Scheme I and Scheme II indicate that the computation is one-dimensional. Second, the invariance of transitions for Scheme I and Scheme II exclude correspondence-channels. This leaves us with a system of multiple energy-channels, that combine into a single scalar representation of motion strength.

Although we have shown that a *single* energy-channel is *sufficient* to model the data, we promised a proof for the *necessity* of a single energy-channel. This proof is based on the inconsistency of multiple energy-channels with transition invariance. We assume a system of multiple energy-channels that *linearly* combine to represent motion strength (summation of energy-channels). Such a system would result in different transitions for Scheme I and II. The proof is given and discussed in the Appendix.

6.2 The Activity Channel

In this section, we derive the characteristics of the single energy-channel. This energy-channel consists of two stages. The first stage is the nonlinear transformation (texture grabber). The simplest version of a texture grabber is a spatiotemporal linear filter followed by rectification (see Chubb and Sperling, 1989). The output of this first stage (the texture activity) is fed into the second stage: standard motion analysis. Stages one and two are sketched in Fig. 8.

— *Figure 8 about here* —

6.2.1 Stage 1: Texture Grabbers

It is now well-established (See review by Shapley and Enroth-Cugell, 1984), that early retinal gain-control mechanisms pass not stimulus luminance, but rather a signal approximating stimulus *contrast*, the normalized deviation of stimulus luminance from its local average. We assume that the spatiotemporal filters of stage 1 operate on stimulus contrast.

The output magnitude of these filters varies over the visual field, depending on what textures happen to populate these regions. The output of a linear filter to a texture is variable and dependents

on the local phase of the texture. The purpose of rectification is to transform regions of highly variable response into regions of high average value, thus insuring that the rectified output registers the presence or absence of texture, independent of phase. Examples of rectification are half-wave rectification (setting negative values to zero) and full-wave rectification (anything that is symmetric with respect to input sign, such as absolute value or squaring).

The output of Stage-1 is called *activity*. The resulting transformation (accomplished by stage 1) yields a spatiotemporal function whose value reflects the local texture preferences of the stage 1 filter in the visual field as a function of time (see also Bergen and Adelson, 1988; Caelli, 1985). The activity transformation of the texture grabber depends on the contrast c and spatial frequency ω of the textures involved.

In Experiment 3, we have shown that texture activity is linear in texture contrast. This is accommodated by a spatial filter that is linear in stimulus contrast. We can further characterize the spatial filter characteristics by the amplitude of its Fourier transform: $F(\omega)$. We assume that rectification is an absolute value operation. Thus, after rectification, the activity transformation T is proportional to c and to $F(\omega)$:

$$T(c,\omega) = cF(\omega)\ . \tag{2}$$

This texture activity T is fed into the second (motion analysis) stage.

6.2.2 Stage 2: Standard Motion Analysis

The second stage (standard motion analysis) is a coincidence detector: it computes the product of the delayed activity at Location 1 with the current activity at Location 2 (van Santen and Sperling, 1984). The output of the second stage corresponds to motion strength.

To simplify the computation in the model, we assume that the first-stage spatiotemporal filter is space-time separable. Indeed, space-time separability seems to be the rule in apparent motion (Burt and Sperling, 1981; van de Grind *et al.*, 1986)[4]. Given space-time separability, we can ignore the temporal component of filtering because temporal patterns were not varied in our stimuli.

We proceed as follows. The perceived direction of motion is considered to be the outcome of a competition in motion strength between motion paths. Within a path the strength of motion between a patch of texture r and a patch of texture s is determined by the product of the activities of the first stage. We assume that the strengths of detectors for all paths are additive in the final motion percept, and adopt a linear combination model (Dosher *et al.*, 1986). Additive internal noise determines the shape of the psychometric functions for motion direction as a function of contrast.

[4]It is reasonable to consider that the linear filter in the texture grabber may itself be composed as a weighted sum of many filters, *i.e.*, filters that also are in the processing path for first-order motion (*e.g.*, Burr *et al.*, 1991). A linear filter composed as the sum of component filters would be space-time separable if each of its component filters were space-time separable and had the same temporal function, independent of spatial scale. This seems to be the case in motion processing (Burt and Sperling, 1981; van de Grind *et al.*, 1986).

Consider the strength model with respect to competition Scheme I (Fig. 1). In one direction there is a homogeneous motion path containing patches of identical texture s. In the opposite direction, there is a heterogeneous motion path containing patches of different textures s and v. For sine wave stimuli, a half-Reichardt model (simple product) is equivalent to the whole Reichardt model (difference of products) (van Santen and Sperling, 1985), so we need to consider just a simple product rule.

The strength of the heterogeneous motion path is:

$$S_{1,he}(c_v, \omega_v, c_s, \omega_s) = c_v F(\omega_v) c_s F(\omega_s). \tag{3}$$

The motion strength $S_{1,ho}$ for the homogeneous motion path is equal to:

$$S_{1,ho}(c_s, \omega_s) = -c_s^2 F^2(\omega_s) \tag{4}$$

(strength in the opposite direction has opposite sign).

Linear combination of both components with equal weights yields a net motion strength D_1 in the direction of the heterogeneous path:

$$D_1(c_v, \omega_v, c_s, \omega_s) = S_{1,he}(c_v, \omega_v, c_s, \omega_s) + S_{1,ho}(c_s, \omega_s). \tag{5}$$

Response variability across trials is due to additive internal noise which is assumed to be distributed as a standard normal density function with mean 0 and standard deviation λ (Fig. 8). A linear addition of noise yields the internal decision variable i which has a normal distribution N with mean D and standard deviation λ.

According to signal detection theory (Green and Swets, 1966) the probability P_1 of heterogeneous motion dominance is:

$$P_1(c_v; \omega_v) = P(i > 0) = \frac{1}{\sqrt{2\pi\lambda^2}} \int_0^\infty N(D_1(c_v, \omega_v, c_s, \omega_s), \lambda)\, di. \tag{6}$$

Substituting motion strengths (Expressions 3 and 4) into the additive linear combination (Expression 5) and then substituting (Expression 5) into the noise-driven decision process (Expr. 6) yields:

$$P_1(c_v; \omega_v) = \frac{1}{\sqrt{2\pi\lambda^2}} \int_0^\infty N([c_v F(\omega_v) c_s F(\omega_s) - c_s^2 F^2(\omega_s)], \lambda)\, di, \tag{7}$$

for the probability of heterogeneous motion dominance for Scheme I (Fig. 1).

Similar reasoning yields the net motion strength D_2 and the probability $P_2(c_v; \omega_v)$ of heterogeneous motion dominance in Scheme II (see Fig. 2):

$$D_2(c_v, \omega_v, c_s, \omega_s) = S_{2,he}(c_v, \omega_v, c_s, \omega_s) + S_{2,ho}(c_s, \omega_s) \quad (8)$$
$$= c_v F(\omega_v) c_s F(\omega_s) - c_v^2 F^2(\omega_v) \quad (9)$$

and

$$P_2(c_v; \omega_v) = \frac{1}{\sqrt{2\pi\lambda^2}} \int_0^\infty N([c_v F(\omega_v) c_s F(\omega_s) - c_v^2 F^2(\omega_v)], \lambda)\, di \quad (10)$$

This model predicts the transition and steepness at transitions of the probability curves for both the experiments.

6.3 Predictions for Scheme I

For different spatial frequencies ω_v of texture v, we measured the probability $P_1(c_v; \omega_v)$ of heterogeneous motion dominance as a function of the contrast c_v of texture v. Our model predicts that the probability P_1 of heterogeneous motion dominance is an error function of the net motion strength D (see Equation 6). In this experiment, the net motion strength D_1 is linear in c_v. Hence, we expect an error function for the probability function $P_1(c_v; \omega_v)$ as a function of c_v (see Equation 7).

6.3.1 Transition Contrast

The transition contrast $\mu_1(\omega_v)$ is defined as the contrast c_v of texture v at which the probability of heterogeneous motion dominance $P_1(c_v; \omega_v)$ is 50% for a given spatial frequency ω_v of texture v. Hence, for $c_v = \mu_1(\omega_v)$, the strength of the heterogeneous and homogeneous motion paths are balanced and we have $S_{1,he} = S_{1,ho}$ or (see Expressions 3 and 4):

$$\mu_1(\omega_v) F(\omega_v) = c_s F(\omega_s) = \kappa, \quad (11)$$

where κ is a constant equal to the activity of standard texture s. texture. If $F(\omega_v)$ is a low-pass filter, $\mu_1(\omega_v)$ will be a monotonically increasing function of ω_v (as supported by our experiments):

$$\mu_1(\omega_v) = \kappa F^{-1}(\omega_s). \quad (12)$$

6.3.2 Steepness

The steepness $\sigma_1(\omega_v)$ is defined as the derivative of $P_1(c_v; \omega_v)$ with respect to c_v at transition contrast $\mu_1(\omega_v)$:

$$\sigma_1(\omega_v) = \frac{\partial}{\partial c_v} P_1(c_v; \omega_v)|_{c_v = \mu_1(\omega_v)} = \frac{1}{\sqrt{2\pi\lambda^2}} \kappa F(\omega_v). \quad (13)$$

Thus, the steepness $\sigma_1(\omega_v)$ is expected to decrease as a function of the spatial frequency ω_v for low-pass filters (as supported by our experiments).

In conclusion we expect error functions for the probability $P_1(c_v;\omega_v)$ of heterogeneous motion dominance as a function of contrast c_v with (a) a transition contrast $\mu_1(\omega_v)$ that is inversely proportional with $F(\omega_v)$ and (b) a steepness $\sigma_1(\omega_v)$ that is proportional with $F(\omega_v)$. If we have low-pass filters, $F(\omega_v)$ decreases monotonically with spatial frequency ω_v.

6.4 Predictions for Scheme II

For different spatial frequencies ω_v of texture v, we measured the probability $P_2(c_v;\omega_v)$ of heterogeneous motion dominance as a function of the contrast c_v of texture v. $P_2(c_v;\omega_v)$ is an error function of D_2 (see Equation 10). However, for Scheme II (unlike for Scheme I) D_2 is not linear with the varied contrast c_v of texture v. As we increase the contrast c_v of texture v, D_2 shows a quadratic dependence on c_v. Therefore, we do not expect an error function for $P_2(c_v;\omega_v)$.

If contrast c_v of texture v is zero, the probability of heterogeneous motion dominance P_2 will be 50% (the motion stimulus is purely ambiguous!). Starting at $c_v = 0$, it first increases linearly with c_v, is maximal for $c_v = c_s F(\omega_s)/[2F(\omega_v)]$, and decreases again with further increases of c_v. Obviously, there may exist a contrast $c_v = \mu_2$ (between the 'optimal' contrast, that yields a maximal D, and a very high contrast, that yields a negative D) for which $P_2 = 50\%$.

Analogous to the derivation in the previous section, one can find the analytic expressions for the transition $\mu_2(\omega_v)$ and steepness $\sigma_2(\omega_v)$ of the probability curves for Scheme II. The expressions for the transition contrasts are equal: $\mu_2(\omega_v) = \mu_1(\omega_v)$. The expressions for the steepness of the transitions for Scheme I and II differ only in sign: $\sigma_2(\omega_v) = -\sigma_1(\omega_v)$.

6.5 The Texture Grabber

We can simply find the Fourier transform $F(\omega)$ of the low-pass filter from the reciprocal transition $\mu_i^{-1}(\omega_v)$ (see Expression 12) and from the steepness $\sigma_i(\omega_v)$ as a function of spatial frequency ω_v (see Expression 13).

The reciprocal transition contrasts are expected to be proportional to the function $F(\omega_v)$. Estimates of the reciprocal transition contrasts $\mu_i^{-1}(\omega_v)$ are shown in Fig. 9.

— Figure 9 about here —

From the reciprocal transitions in Fig. 9, it follows that $F(\omega)$ is a low-pass filter in the range of frequencies examined.

The model predicts that the steepness of the probability function is proportional with the function $F(\omega_v)$ and inversely proportional with λ (the strength of the internal noise). Thus, unlike the transition contrast, the steepness is biased by the internal noise contribution. If the relative strength is constant and independent of the spatial frequency and contrast of the patches of texture involved, the steepness

$\sigma_i(\omega_v)$ is expected to be proportional with $F(\omega_v)$. Estimates of $\sigma_i(\omega_v)$ are shown in Fig. 6. The steepness shows a tendency to decrease with increasing spatial frequency. However, we find some non-monotonicity, in particular for higher spatial frequencies. This may reflect a certain variability of the internal noise for different spatial frequencies.

7 Experiment 4: Perceived Contrast

We have discussed texture grabbers and motion analysis in terms of objective contrast of patches of texture. The experiments implied that the activity of the texture grabber increases monotonically with objective contrast and decreases monotonically with spatial frequency. An interesting question is whether this relation is consistent with the subjective contrast of static gratings as a function of spatial frequency. In other words, is the activity of a texture grabber simply proportional to the subjective contrast?

To answer this question, we performed a contrast discrimination experiment.

7.1 Method

In a two interval presentation subjects looked at an annulus containing either gratings s or v. In one interval we showed an annulus of gratings s (see frame f_2 of Fig. 1), with fixed contrast $c_s = 0.5$ and fixed spatial frequency $\omega_s = 4.9$ cpd. In the other interval we showed an annulus of gratings v (see frame f_2 of Fig. 2), with contrast c_v and spatial frequency ω_v. The order of presentation of the intervals was randomized. Each annulus was shown for 133 ms (which is equal to the frame display time in the motion stimulus). The intervals were separated by a time interval of 133 ms in which the screen was uniform with background luminance. Apparatus, viewing conditions, and other aspects were identical to the motion experiment

7.2 Procedure

The task of the subject was to indicate the interval that contained the patches of grating with the highest contrast. We measured the probability $P_c(c_v; \omega_v)$ that observers judge the grating v as the grating with the highest contrast as a function of the objective contrast c_v of grating v. In the contrast matching experiment, we examined two spatial frequencies: $\omega_v = 1.2$ cpd, and $\omega_v = 7.4$ cpd of grating v. These were the lowest and highest spatial frequencies for which we found transition invariance in our motion experiment. From these probability curves, we estimated the *matching contrast* of grating v for which the perceived contrast of grating s and v was equal. The precise estimation of the matching contrast was analogous to the estimation of transition contrast in the motion competition experiments.

7.3 Results

— Figure 10 about here —

In Fig. 10, we show the probabilities of judging the contrast of grating v higher than that of grating s (with $c_s = 0.5$) as a function of objective contrast c_v (filled circles). For all conditions and subjects, the perceived contrast of texture v increases monotonically with its objective contrast c_v. The contrast c_v where the curve crosses the 50% guide line is the *matching contrast*. For a 'low' spatial frequency grating v ($\omega_v = 1.2$ cpd), we find that the perceived contrasts of s and v are matched when $c_v = 0.47$ for subject PW and $c_v = 0.44$ for JS. This matching contrast is close to the objective contrast $c_s = 0.5$ of grating s. For a 'high' spatial frequency grating v ($\omega_v = 7.4$ cpd), the matching contrasts are $c_v = 0.54$ for PW and $c_v = 0.53$ for JS.

For comparison of the matching contrast with the transition contrast in the motion experiments, we have also shown the probabilities to perceive heterogeneous motion using Scheme I as a function of c_v in the corresponding panels.

7.4 Discussion

Interestingly, the matching contrasts for low and high spatial frequency gratings are approximately equal to the objective contrast of grating s, for the range of contrasts and spatial frequencies of grating v examined. That is, perceived contrast does not depend on spatial frequency. However, the contrast of grating v for balancing the motion paths when $\omega_v = 1.2$ cpd for Scheme I was: $c_v = 0.22$ for subject PW and $c_v = 0.36$ for JS. Obviously, at the transition contrast for the motion experiment, the perceived contrast of grating s and v are markedly different. That is, the activities of the grating v are matched even when *both* spatial frequency and perceived contrast are different from grating s. In conclusion, activity can not be a function that depends solely on perceived contrast.

8 Experiment 5: Dichoptic Presentations

8.1 Motivation

We have succesfully modeled the strength of motion-from-texture in terms of a texture grabber followed by standard motion analysis. Standard motion analysis is a type of motion computation that is *not* sensitive to correspondences in textural features. An interesting property of standard motion analysis is that the neural substrate for such a process is organised so as to require successive stimulation to the same eye. When monocular motion information is not available to the observer standard motion analysis fails.

The motion system that extracts information of both eyes (when motion is presented dichoptically) can be classified as a *correspondence-channel*. For example, Pantle and Picciano (1976) studied apparent motion with a three dot stimulus and reported element movement for monocular and binocular presentation, but group movement for dichoptic presentation. The group movement suggests a representation of features or shapes precedes the extraction of motion. Also, Georgeson and Shackleton (1989) showed that drifting square-wave gratings with missing fundamental (MF) moved backwards while presented monocularly (following the third harmonic) but moved forwards when presented dichoptically. They suggested that the perceived direction of dichoptic apparent motion was consistent

with a system that combines information across spatial frequency channels to identify local features and then tracks the location of corresponding features over time.

Following the above reasoning, the motion system for dichoptic presentations would be sensitive to the similarity of the textures involved. Thus, the contribution of what we call correspondence-channels might be more pronounced when our competition schemes are presented dichoptically (sofar viewing has been binocular in our experiments). We tested our energy-channel model for both dichoptical and monocular presentations of our motion stimuli. This test may also locate the motion extraction process involved in our stimuli in terms of different levels in the visual nervous system (before or after the sites of binocular combination).

8.2 Results

The ambiguous motion competition schemes I and II can be presented dichoptically in two different modes. In the first mode, the odd frames are presented in one eye and the even frames in the other. In this way, the spatiotemporal stimulus is purely ambiguous in each eye. Both the heterogeneous and the homogeneous paths are processed by dichoptic mechanisms. In this mode, dichoptic mechanisms are not competing with monocular mechanisms.

In the second mode, the patches of one texture type are presented in one eye and the patches of the second type of texture in the other eye. In this way the homogeneous motion path (textures s for Scheme I) is presented in one eye, while the textures v in the other eye form a purely ambiguous stimulus. In this mode, dichoptic mechanisms processing the heterogeneous path have to compete with monocular mechanisms processing the homogeneous path.

We determined the psychometric functions for both competition schemes for a condition where the texture s and v differ two octaves in spatial frequency ($\omega_s = 4.9$ cpd and $\omega_v = 1.2$ cpd) for subject PW. The binocular results were presented in top-left panel of Fig. 4. As discussed for Experiment 1 and 2, a difference between the transition contrasts μ_1 and μ_2 indicates the involvement of additional (correspondence) channels. The results for *monocular* presentation were identical (within measurement error) to the results for *binocular* presentation. For both conditions, we find transition invariance: $\mu_1 = \mu_2 \approx 0.2$.

The results for both modes of *dichoptic* presentation were very similar to those for *binocular* presentation. That is, dichoptic presentation yields psychometric functions for Scheme I and II similar to those for binocular presentation. For adequate contrast c_v heterogeneous motion dominated homogeneous motion for both modes of dichoptic presentation suggesting the dominance of an energy-channel even when monocular motion information was absent. However, the contribution of a correspondence-channel is noticable for dicoptic presentations, transition invariance no longer holds. We found $\mu_1 \approx 0.2$ and $\mu_2 \approx 0.1$ for both modes of dichoptic presentation.

8.3 Discussion

Motion perception between patches of non-similar texture is easily perceived for both modes of dichoptic presentation (as predicted by our energy-channel). Even in the second mode, where a dichoptic

heterogeneous motion path competes with a monocular homogeneous path, heterogeneous motion can easily dominate for small contrast of texture v (*e.g.*, $c_v > 0.2$ for Scheme I). These results suggest that dichoptic processing of our motion stimuli is dominated by the same mechanisms as monocular processing and that motion strength is not predicted by the similarity between textural features such as spatial frequency.

However, although dichoptic presentation leaves transition contrast μ_1 for Scheme I unaffected, transition μ_2 for Scheme II decreases. This difference from the binocular results indicates a significant contribution of other channels when monocular information for the heterogeneous path is ambiguous. A more detailed investigation might be useful.

9 General Discussion

9.1 Fallacy of Correspondence Matching

The experiments presented in this paper provide cogent evidence that texture similarity is not relevant to the texture-from-motion computation (within the range of spatiotemporal parameters varied in this experiment). As an example it was shown that motion between patches of texture that differ by two octaves in spatial frequency and a factor of two in contrast can be stronger than motion between patches of identical texture.

The correspondence matching metaphore to explain visual processes in several visual domains seems to have lost predictive power. Correspondence matching fails to explain the dominance of (1) heterogeneous motion paths composed of textures that differ in spatial frequency and contrast (this paper), (2) heterogeneous motion paths composed of elements that differ in size, orientation and luminance (Werkhoven *et al.*, 1990a,1990b), and (3) stereoscopic matches between elements that differ in size and luminance (Gulick and Lawson, 1976).

The visual motion system does not seem to be designed to establish correspondence between similar features in a motion sequence. This should not come as a surprise given the inherent difficulties in designing correspondence matching mechanisms. Such mechanisms would look for 'similar features' in 'successive' time samples of the spatiotemporal stimulus. However, what constitutes a feature, and how strict should similarity be taken?

Recently developed stimulus (motion) energy models for motion extraction bypass the correspondence problem and are more likely candidates for the kind of visual processing early in the visual system (Adelson and Bergen, 1985; Heeger, 1992). The energy-channel described in this paper is equivalent to such an motion energy computation, applied to a nonlinear transformation of the stimulus, (van Santen and Sperling, 1984).

9.2 Contrast and Motion

In Experiment 3, we showed that the transition contrast of texture v needed to balance the motion path $s-v$ with the motion path $s-s$ varies linearly with the contrast of texture s. In the context of our model, this means that the activity of a texture grabber is approximately linear in texture contrast.

In fact, we find linearity even for high contrasts in the range of 50% to 100%. As a consequence of this contrast linearity, motion strength varies linear with the contrast of each of the texture inputs. That is, the strength of motion between two textures with identical texture contrast is quadratic with this contrast. Approximate contrast linearity of the input lines for standard motion analysis was also found for experiments with spatiotemporal modulations of luminance Werkhoven *et al.* (1990b).

It should be noted, that the linear contrast dependency is at odds with the contrast thresholds for motion direction discrimination reported by Nakayama and Silverman (1985). They measured the smallest phase shift (yielding threshold direction discrimination performance) of sinusoidal gratings as a function of grating contrast. The smallest phase shift yielding threshold performance leveled off for grating contrasts exceeding 5%. They interpreted their finding in terms of a contrast saturation function. However, their results are open to a different interpretation in which the minimum phase shift is limited by other (spatial) properties of the motion extraction mechanism leaving the contrast dependency unknown.

9.3 A Shared Motion Analysis Stage?

An intriguing question is how mechanisms for the extraction of motion carried by the spatiotemporal modulation of luminance relate to those for extracting motion carried by the spatiotemporal modulation of texture type. To discriminate both mechanisms we have to compare the characteristics of the perception of both motion types. For example, Turano and Pantle (1989) studied velocity discrimination performance for both types of motion stimuli and showed similar discrimination characteristics. Their results support the hypothesis of a higher order (motion analysis) mechanism that accepts input from both the luminance-domain as well as texture-domain.

A shared motion analysis stage for the two types of motion is also supported by our finding that strength of motion-from-texture is ruled by the same metric as motion in the luminance domain. Motion strength is the covariance (or product) of local activities. This activity is simply the luminance itself when the motion is carried by luminance (van Santen and Sperling, 1985) or a nonlinear transformation of the luminance pattern for motion-from-texture (this paper).

In conclusion, the extraction of motion from the spatiotemporal modulations of luminance and that of texture type seems to be mediated by a shared standard motion analysis stage. However, additional experiments with different paradigms may weaken this idea. For example, Mather (1991) showed that both motion types produce motion after effects, but that the duration of the aftereffects were significantly different.

9.4 Transitivity and Additivity

Under the assumption of standard motion analysis and channel summation, the metamery of motion paths $s - v$ showed in this paper implies that the corresponding patches of texture v are metamer with s with respect to motion processing. That is, all textures v of this metameric class yield identical motion strength when embedded in a motion path $s - v$.

Metamery yields two strong predictions. First, metamery predicts *transitivity*: if textures a and

b are metamer with s, then a is metamer with b. Second, metamery predicts *additivity*: if textures a and b are metamer with s, then any linear combination $\alpha a + \beta b$ (with $\alpha + \beta = 1$) is metamer with s.

These predictions have not yet been tested.

9.5 Motion Transparency

The energy-channel proposed in this paper computes the difference between left and rightward motion. This implies that motion transparency (the simultaneous detection of left and rightward motion) is not readily accommodated in this model. Because the motion analysis component of the energy-channel is a Reichardt-correlator, the motion energy of the left and rightward motion path are no explicit intermediate results). However, occasionally, observers reported transparency for stimuli that were nearly balanced.

Adelson and Bergen (1985) addressed this issue by pointing out that although their energy detector was functionally equivalent to correlation detector, the intermediate results are not. Specifically, the energy of left and rightward motion are explicit intermediate results in energy detectors, but not in correlation detectors (the output of a half Reichardt-correlator is the half-phase opponent energy!). Although our conclusions do not depend on the specific choice of motion model, a further study of transparency in this context might reveal the specific type of detector involved.

9.6 Extension of the Parameter Space

It is important to remember that we have shown the one-dimensionality of the motion-from-texture computation only with respect to parallel sinewave patches that differ in spatial frequency and contrast. Chubb and Sperling (1991) found that motion-from-texture could be carried by differences in spatial orientation, although differences in orientation did not produce as vigorous motion as did differences in spatial frequency. This observation indicates that orientation (and possibly other properties) are relevant to motion-from-texture. It would be interesting to determine the dimensionality of the computation for a larger class of stimuli.

Although motion strength at a 'frame time' τ of 8/60 sec is exclusively determined by the product of activities, we can not exclude that effects of texture similarity are stronger at longer frame time. In fact, the temporal frequency of texture modulation in our experiments is 1.9 Hz (one cycle consists of four frames of 133 ms each). At slower temporal frequencies, the processing time for the textures increases, perhaps enabling more elaborate 'texture grabber' filters or correspondence-channels to contribute to motion strength.

Effects of other properties (*e.g.*, orientation) and temporal parameters are currently under investigation.

10 Acknowledgement

This work was supported by the USAF Life Science Directorate, Visual Information Processing Grant 88-0140.

References

[1] Anstis S.M. (1980) The perception of apparent movement, *Philosophical Transactions of the Royal Society, London* **B 290**, 153–168.

[2] Adelson E.H. and Bergen J.R. (1985) Spatio-temporal energy models for the perception of motion, *J. of the Optical Society America A* **2**, No. **2**, 284–299.

[3] Balliet R. and Nakayama K. (1978) Training of voluntary torsion, *Investigative Ophthalmology & Visual Science* **17**, No. **4**, 303–314.

[4] Bergen J.R. and Adelson E.H. (1988) Early vision and texture perception, *Nature* **333**, 363–364.

[5] Braddick O.J. (1980) Low-level and high-level processes in apparent movement, *Philosophical Transactions of the Royal Society, London* **B 290**, 137–151.

[6] Burt P. and Sperling G. (1981) Time, distance and feature trade-offs in visual apparent motion, *Psychological Review* **88**, No. **2**, 171–195.

[7] Caelli T. (1985) Three processing characteristics of visual texture segregation, *Spatial Vision* **1**, No. **1**, 19–30.

[8] Cavanagh P. and Mather G. (1989) Motion: The long and the short of it, *Spatial Vision* **4**, No. -, 103-129.

[9] Chubb C. and Sperling G. (1988) Drift-balanced random stimuli: a general basis for studying non-Fourier motion perception, *J. of the Optical Society America A* **5**, No. **11**, 1986–2007.

[10] Chubb C. and Sperling G. (1989) Second-Order Motion Perception: Space/time Separable Mechanisms, *Proceedings: Workshop on Visual Motion, Irvine, California, 1989.* IEEE Computer Society Press, pp. 126–138, 1989.

[11] Chubb C. and Sperling G. (1991) ms in press, 1991

[12] Dosher B.A., Sperling G and Wurst S.A. (1986) Tradeoffs between stereopsis and proximity luminance covariance as determinants of perceived 3D structure, *Vision Research* **26**, No. **6**, 973–990.

[13] Duda R.O. and Hart P.E. (1973) Pattern classification and Scene Analysis, Wiley-Interscience, 1973.

[14] Georgeson M.A. and Shackleton T.M. (1989) Monocular motion sensing, binocular motion perception, *Vision Research* **29**, No. **11**, 1511–1523.

[15] Green M. (1986) What determines correspondence strength in apparent motion?, *Vision Research* **26**, No. **4**, 599–607.

[16] Green D.M. and J.A. Swets (1966) Signal detection theory and psychophysics, New York: Wiley, 1966.

[17] Grind W.A. van de, Koenderink J.J. and Doorn A.J. van (1986) The distribution of human motion detector properties in the monocular visual field, *Vision Research* **26**, No. **5**, 797–810.

[18] Gulick W.L. and Lawson R.B. (1976) Human stereopsis: a psychophysical analysis, Oxford University Press, New York, 1976.

[19] Farrell J.E., Pavel M. and Sperling G. (1990) The visible persistence of stimuli in stroboscopic motion, *Vision Research* **30**, No. **6**, 921–936.

[20] Graham N. (1992) Complex channels, early local nonlinearities, and normalization in texture segregation. In: Computational Models of Visual Processing, Chapter 18. Edited by M. S. Landy and J.A. Movshon. MIT Press, 1992.

[21] Heeger D.J. (1992) Nonlinear model of neural responses in cat visual cortex. In: Computational Models of Visual Processing, Chapter 9. Edited by M. S. Landy and J.A. Movshon. MIT Press, 1992.

[22] Helmholtz H. von (1924) Physiological Optics, Vol. II, Trans. J. Southall. Rochester, N.Y.: Optical Society of America.

[23] Kolers P.A. (1972) Aspects of Motion Perception, Pergamon Press Oxford.

[24] Lelkens A.M.M. and Koenderink J.J. (1984) Illusory motion in visual displays, *Vision Research* **24**, No. **9**, 1083–1090.

[25] Mather G. (1991) First-order and second-order visual processes in the perception of motion and tilt, *Vision Research* **31**, No. **1**, 161–167.

[26] Nakayama K. and Silverman G.H. (1985) Detection and discrimination of sinusoidal grating displacements, *J. of the Optical Society America A* **2**, No. **2**, 267–274.

[27] Navon D. (1976) Irrelevance of figural identity for resolving ambiguities in apparent motion, *J. of Experimental Psychology, Human Perception and Performance* **2**, No. **1**, 130–138.

[28] Pantle A. and Picciano L. (1976) A multistable movement display: Evidence for two separate motion systems in human vision, *Science* **193**, 500–502.

[29] Papathomas T.V., Gorea A. and Julesz B. (Two carriers for motion perception: color and luminance) *Vision Research* ,, No. **31**, 11.1991

[30] Reichardt W. (1961) Autocorrelation, a principle for the evaluation of sensory information by the central nervous system, In *Sensory Communication*, W.A. Rosenblith, ed. (Wiley, New York, 1961).

[31] Richards W. (1979) Quantifying Sensory Channels: Generalizing Colorimetry to Orientation and Texture, Touch, and Tones, *Sensory Processes* **3**, 207–229.

[32] Santen J.P.H. van and Sperling G. (1985) Elaborated Reichardt detectors, *J. of the Optical Society America A* **2**, No. **2**, 300–321.

[33] Shapley R. and Enroth-Cugell C. (1984) Visual adaptation and retinal gain controls, *Prog. Retinal Res.*, **B3**, pp. 263-346, 1984.

[34] Shechter S., Hochstein S. and Hillman P. (1989 a) Size, flux and luminance effects in the apparent motion correspondence process, *Vision Research* **29**, No. **5**, 579–591.

[35] Sperling G. (1976) Movement perception in computer-driven visual displays, *Behavior Research Methods & Instrumentation* **8**, No. **2**, 144–151.

[36] turano K. and Pantle A. (1989) On the mechanism that encodes the movement of contrast variations: velocity discrimination, *Vision Research* **29**, No. **2**, 207–221.

[37] Ullman S. (1980) The effect of similarity between bar segments on the correspondence strength in apparent motion, *Perception* **9**, No. **617–626**, 1980.

[38] Victor J.D. and Conte M.M. (1990) Motion mechanisms have only limited access to form information, *Vision Research* **30**, No. **2**, 289–301.

[39] Watson A.B. (1986) Apparent motion occurs only between similar spatial frequencies, *Vision Research* **26**, No. **10**, 1727–1730.

[40] Werkhoven P., Snippe H.P. and Koenderink J.J. (1990) Effects of Element Orientation on Apparent Motion Perception, *Perception & Psychophysics* **47**, No. **6**, 509–525.

[41] Werkhoven P., Snippe H.P. and Koenderink J.J. (1990) Metrics for the Strength of Low Level Motion Perception, *J. of Visual Communication and Image Representation* **1**, No. **2**, 176–188.

[42] Werkhoven P. and Koenderink J.J. (1991) Reversed Rotary Motion Perception, *J. of the Optical Society America A* **8**, No. **9**, 1510–1516.

[43] Williams D.W., Tweten S. and Sekuler R. (1991) Using metamers to explore motion perception, *Vision Research* **31**, No. **2**, 275–286.

11 Appendix: Multiple Activity Channels and Transition Invariance

11.1 A System of Multiple Activity-channels

We propose a *multi-channel* model (multiple energy-channels) for computing the strength of motion-from-texture. The model consists of two stages, as shown in Fig. 11.

— Figure 11 about here —

11.1.1 Stimulus Transformation: Texture Grabbers

Stage 1 consists of n types of texture grabbers—where each type of texture grabber i is described by nonlinear spatiotemporal transformations T_i, $i = 1 \ldots n$, of the optical input. Each transformation yields a spatiotemporal function $T_i(\varphi, t)$ whose value reflects the local texture preferences of the Stage 1 filters in the visual field as a function of position φ and time t. (We use φ for the position of a texture grabber because, in our essentially one-dimensional stimulus, the texture position is determined by the angle φ.) The output of these texture grabbers is called *activity*. The n different transformations T_i of Stage 1 transform the optical input into n *activity representations*.

11.1.2 Motion Detection

Stage 2 is a set of motion detectors. For specificity, but without loss of generality (see van Santen and Sperling, 1985; Chubb and Sperling, 1988, 1991) we adopt Reichardt's scheme for standard motion analysis (Reichardt, 1961) which consists of two oppositely tuned coincidence detectors. Motion detectors operate on the outputs of the texture grabbers. Each type of texture grabber (transformation T_i) has its own, unique set of motion detectors. A transformation T_i together with its motion detectors is called a motion channel i).

A coincidence detector performs a multiplication operation on the current activity $T_i(\varphi, t)$ at position φ at time t and the (delayed) activity $T_i(\varphi - \Delta\varphi, t - \Delta t)$ at position $\varphi - \Delta\varphi$ and time $t - \Delta t$. Hence, the output of the coincidence detector is: $T_i(\varphi - \Delta\varphi, t - \Delta t)T_i(\varphi, t)$. The outputs of two coincidence detectors tuned to identical velocities but opposite directions are subtracted to yield a net motion strength $D_i(\varphi, t)$:

$$D_i(\varphi, t) = T_i(\varphi - \Delta\varphi, t - \Delta t)T_i(\varphi, t) - T_i(\varphi - \Delta\varphi, t)T_i(\varphi, t - \Delta t). \quad (14)$$

Channel i has a positive output for motion in the direction of positive φ and a negative output for motion in the opposite direction.

11.1.3 Summation

In a one-dimensional motion computation, the outputs of a system of energy-channels described above (represented in a n dimensional channel space) are essentially mapped to a single (decision) dimension: the final net motion strength. This mapping maps a $n-1$ dimensional manifold in the channel space to a single point in the one-dimensional decision space (final motion strength). For example, channel summation maps a planar surface in the channel space to zero final motion strength (for Scheme I). For other combination rules than summation, other (non-planar) surfaces will map to zero final motion strength. However, when we assume that this mapping is *continuous* and *differentiable*, these true manifolds are in first order approximated by a planar surface for *small* channel signals at transition points. Channel summation is a sufficient first order combination rule.

Summation of channels D_i yields net motion strength D:

$$D(\varphi,t) = \sum_{i=1}^{n} D_i(\varphi,t). \tag{15}$$

11.2 Predictions for Competition Schemes

We apply the *multi-channel* computation to competition schemes I and II (see Fig. 1 and Fig. 2. Consider first Scheme I. The heterogeneous path is the motion between texture s (at time $t-\Delta t$ and position $\varphi-\Delta\varphi$) and texture v (at time t and position φ). Let $T_{i,s}$ be the activity of texture grabber T_i for texture s, and $T_{i,v}$ the activity of texture grabber T_i for texture v. The output of channel i for this path is the product of the delayed activity $T_{i,s}$ of texture s and the current activity $T_{i,v}$ of texture v. For simplicity, we will use the vector notation:

$$\vec{T_s} = \begin{pmatrix} T_{1,s} \\ T_{2,s} \\ . \\ . \\ T_{n,s} \end{pmatrix} \quad \text{and} \quad \vec{T_v} = \begin{pmatrix} T_{1,v} \\ T_{2,v} \\ . \\ . \\ T_{n,v} \end{pmatrix} \tag{16}$$

The vectors $\vec{T_s}$ and $\vec{T_v}$ are the activity vectors of textures s and v respectively. An activity vector represents the activity of a texture in the n-dimensional transformation space (T-space) defined by transformations $T_1 \ldots T_n$.

For Scheme I, the motion strengths $S_{1,he}$ summed over all channels for the heterogeneous path can be written as the vector product:

$$S_{1,he} = \vec{T_s} \cdot \vec{T_v} = \sum_{i=1}^{n} T_{i,s} T_{i,v}. \tag{17}$$

We have arbitrarily assigned a positive sign to motion strength in this direction. Motion in the opposite direction has a negative sign (see Equation 14). The output of channel i for the homogeneous path (between textures s) is the squared output of transformation $T_{i,s}$. The motion strength S_{ho} of the homogeneous path is (after summing all channels) is:

$$S_{1,ho} = -\vec{T}_s \cdot \vec{T}_s. \tag{18}$$

Adding Equations (6) and (7) gives the net motion strength D_1 in the direction of the heterogeneous path for Scheme I:

$$D_1 = \vec{T}_s \cdot [\vec{T}_v - \vec{T}_s]. \tag{19}$$

Analogously, the net motion strength D_2 in the direction of the heterogeneous path for Scheme II is:

$$D_2 = \vec{T}_v \cdot [\vec{T}_s - \vec{T}_v]. \tag{20}$$

11.3 Transitions: Scheme I

At a transition for Scheme I, the net motion strength D_1 is zero:

$$D_1 = \vec{T}_s \cdot [\vec{T}_v - \vec{T}_s] = 0. \tag{21}$$

There exists an $(n-1)$-dimensional plane of $\vec{T}_v$ vectors in T-space for which the motion strength of the heterogeneous and homogeneous motion paths are balanced (the vectors $\vec{T}_v$ for which the difference vector $\vec{T}_v - \vec{T}_s$ are orthogonal to vector $\vec{T}_s$).

— Figure 12 about here —

In fact, the solution space is even more constrained than shown in Fig. 12a. Let each texture grabber be a function of m textural properties. If we consider the m-parameter space that characterize our textures and an n-dimensional T-space, than the parameter space is mapped on a m-dimensional surface in T-space. Possible solutions are the intersections of this surface with the solution plane.

Consider, for example, a two-dimensional T-space (a two-channel motion computation). The vectors $\vec{T}_v$ in T-space that satisfy Equation 21 for a certain vector $\vec{T}_s$ must end on the dashed guide line in Fig. 12a. Unless all transformations T_i are identical, each vector $\vec{T}_v$ of this solution will project back to a unique point (texture) in our parameter space. Thus, the activity vectors that yield

balanced motion strength for a particular texture s, are described by a curve in the parameter space (*e.g.*, frequency/contrast space in our experiments).

It should be noted in passing, that the net heterogeneous motion strength $D_1 = \vec{T}_s \cdot [\vec{T}_v - \vec{T}_s]$ can be positive. Hence, even in a multi-channel computation, the strength of the heterogeneous motion path can dominate.

11.4 Transitions: Scheme II

Similarly, at a transition for Scheme II (Fig. 2), the net motion strength D_2 is zero:

$$D_2 = \vec{T}_v \cdot [\vec{T}_s - \vec{T}_v] = 0. \tag{22}$$

The $(n-1)$-dimensional solution of $\vec{T}_v$ vectors in T-space for which the motion strength of the heterogeneous and homogeneous motion paths are balanced is not a plane. For example, we consider again the two-dimensional T-space. The vectors $\vec{T}_v$ in T-space that satisfy Equation 22 for a certain vector $\vec{T}_s$ end on a circle containing $\vec{T}_s$ (see Fig. 12b). Again we will find a one-dimensional solution in the parameter space. However, it will differ from the solution for Scheme I, when T-space is two-dimensional (or higher dimensional).

11.5 Transition Invariance

Using only the result for Scheme I, we cannot discriminate between a *single-channel* ($n = 1$) and *multi-channel* computations ($n > 1$): either single- or multi-channel computations might yield solutions to Equation 21. To resolve the issue, we need the constraint of transition invariance.

Transition invariance means that once the motion strength of the heterogeneous path and that of the homogeneous motion path are balanced for a particular pair of textures s and v for Scheme I, this balance is not disturbed by interchanging the textures s and v (yielding Scheme II). We now show that transition invariance is inconsistent with a *multi-channel* computation.

The transitions are invariant iff the activity vector $\vec{T}_v$ simultaneously satisfies Equations 21 and 22. Because the difference vector $\vec{T}_s - \vec{T}_v$ is always in the plane defined by vector $\vec{T}_s$ and vector $\vec{T}_v$, the only vector $\vec{T}_v$ that satisfies both equations is $\vec{T}_v = \vec{T}_s$.

Vector $\vec{T}_v$ is equal to vector $\vec{T}_s$ iff each transformation T_i involved in the motion computation has an equal output for both textures v and s:

$$T_{i,s} = T_{i,v} \quad (i = 1 \ldots n). \tag{23}$$

Equation 23 represents a very strong constraint for the ensemble of transformations that might be involved in a *multi-channel* computation. Every transformation T_i must have an iso-activity contour as a function of all textural properties (*e.g.*, frequency-contrast space) that contains both the activity of texture s and that of texture v. Furthermore, transition invariance holds for different texture pairs (s, v); the iso-activity contours of each transformation T_i must be identical for all these pairs.

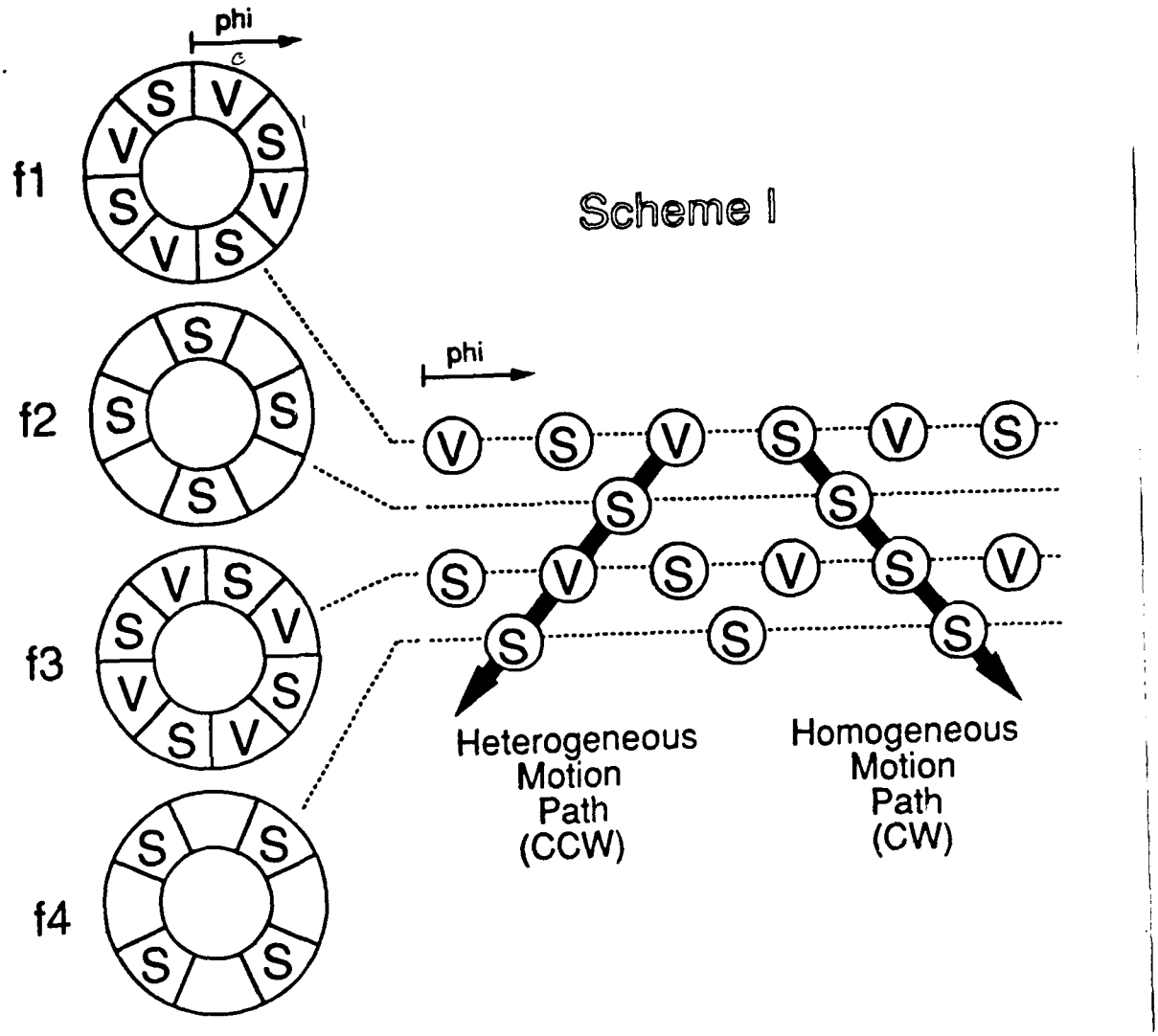


Figure 1: Motion competition Scheme I. *At the left side:* A series of frames ($f_1, f_2, \ldots$) is shown successively in time (for details see Section 'Method'). The first frame (f_1) contains an annulus of patches of alternated texture type s and v at regular positions drawn against a uniform background. The annulus has an inner radius of $r_1 = 1.04$ degrees of visual angle, and an outer radius of $r_2 = 2.08$ deg. The patches of texture s and texture v are spatially contiguous and alternate within the annulus. Since the annulus contains 8 patches, each patch has a width of 45 degrees. Angular position φ is measured clock-wise with respect to the vertical.
The second frame (f_2) is similar to frame f_1, except that the low frequent patches of texture v are now replaced by a uniform patch of background luminance. Furthermore, f_2 is rotated (clockwise) around the center of the annulus over an angle of 22.5 degrees with respect to frame f_1. In a sequence of frames, frame f_{n+2} is identical to frame f_n , except for a rotation around the center over an angle of 45 degrees (clockwise).
At the right side: Angular positions φ is along the horizontal axis. Patches of texture s and v are shown at their angular positions for frames $f_1 \ldots f_4$ yielding rows of patches. The top row of patches s and v corresponds to frame f_1. The second row of patches s corresponds to frame f_2. Hence, time (or frame number) is along the vertical axis.
When frame f_n and frame f_{n+1} are presented in succession, two motion paths are *a priori* likely. A homogeneous motion path: clockwise matches (CW) between patches of identical texture s (indicated by the arrow pointing down and right). A heterogeneous motion path: counter-clockwise (CCW) matches between patches of texture s and patches of texture v (indicated by the arrow pointing down and left).

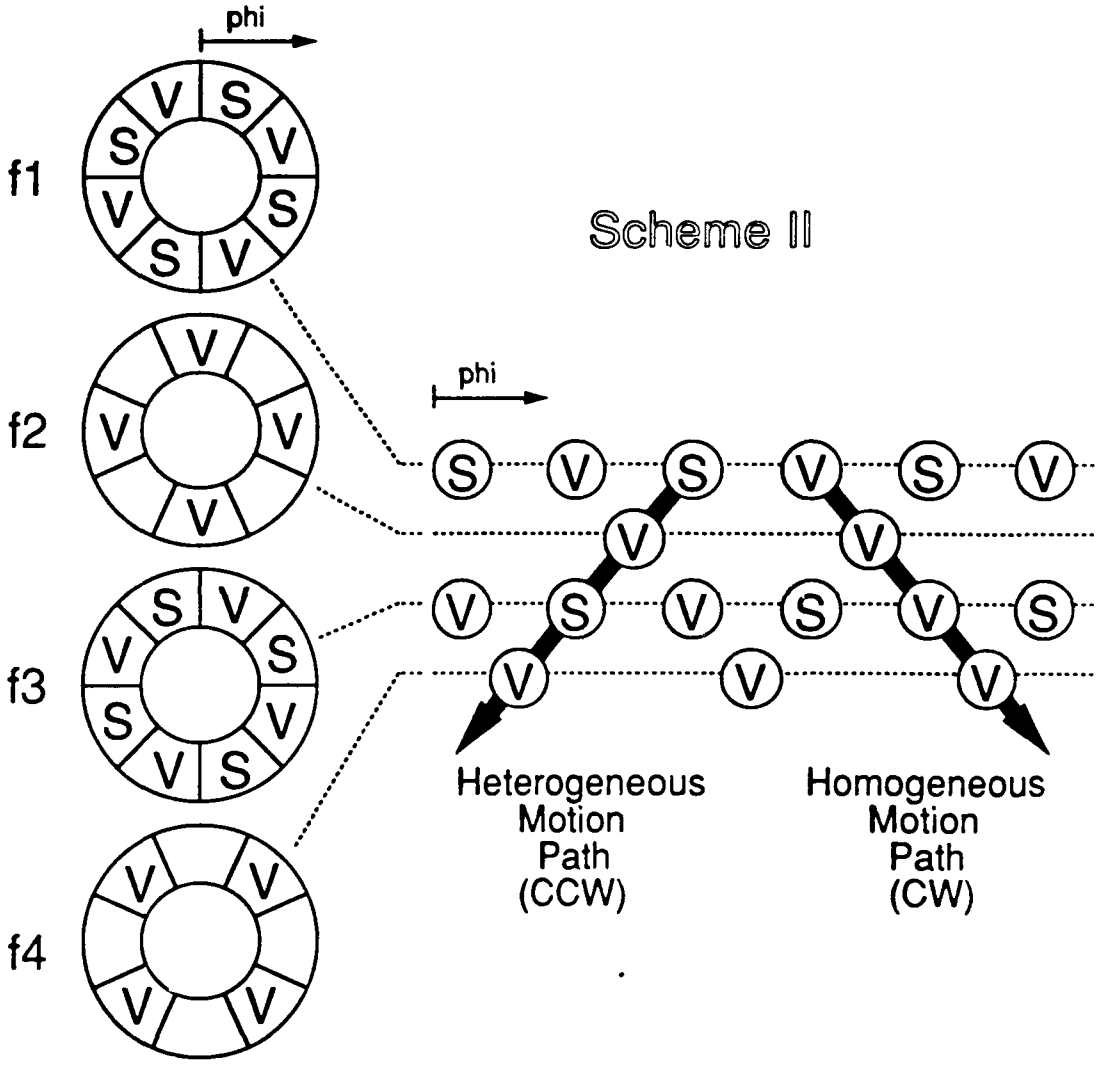


Figure 2: Motion competition Scheme II. This scheme is similar to Scheme I (see Fig. 1), except that textures s and v are interchanged. In Scheme II, the homogeneous motion path contains textures v.

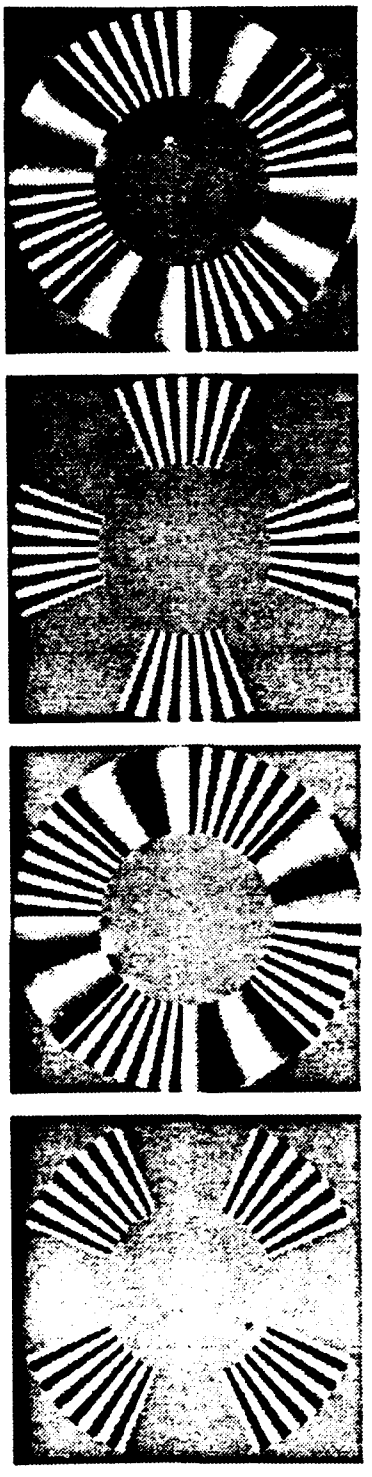


Figure 3: An example of the ambiguous motion display (as sketched in Fig. 1). Frames f_1, f_2, f_3, and f_4 (containing the patches of textures) are shown in (a), (b), (c) and (d) respectively. For this example, textures s and v differ only in their spatial frequency: the spatial frequency of texture s is two octaves higher than that of texture v.

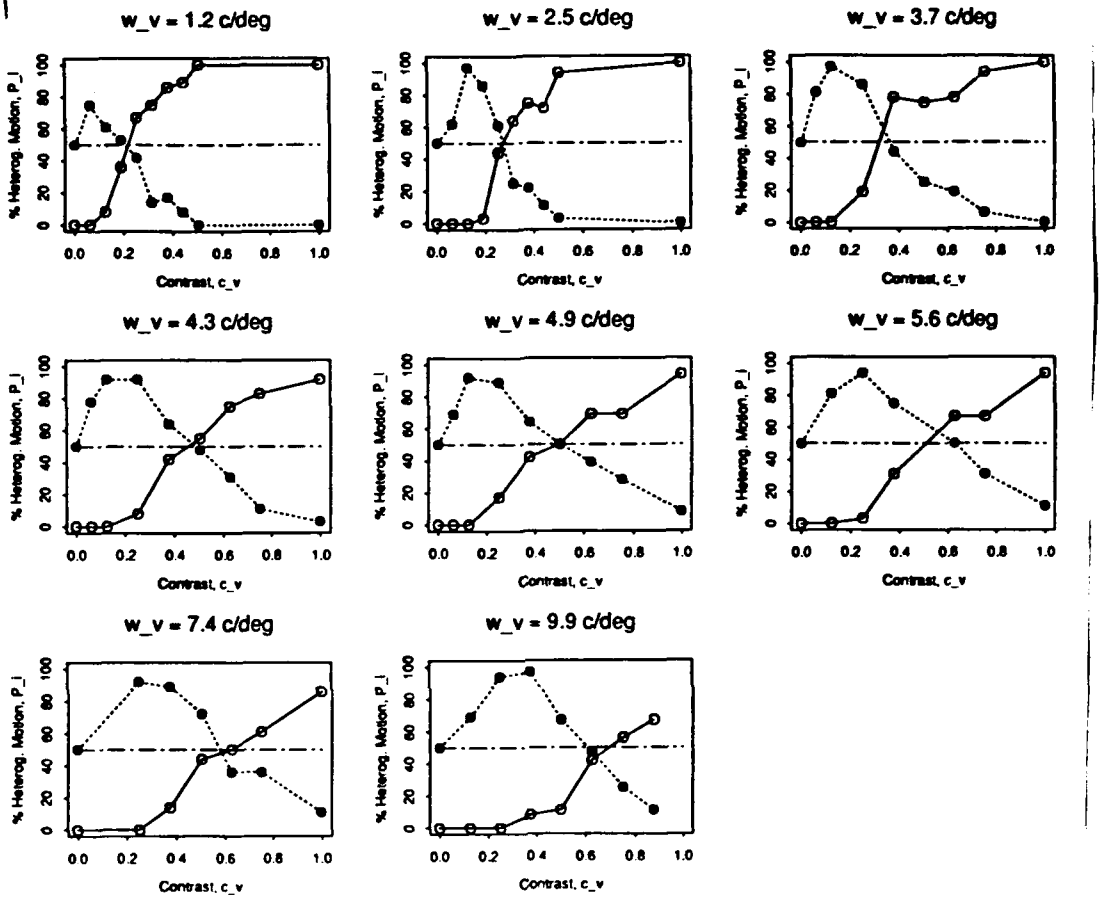


Figure 4: Probability $P_i(c_v;\omega_v)$ of the dominance of a heterogeneous motion path over a homogeneous motion path is shown as a function of the contrast c_v of texture v for different spatial frequencies ω_v of texture v for two subjects. *Open circles* represent the probability $P_1(c_v;\omega_v)$ for Scheme I (Fig. 1), *filled circles* $P_2(c_v;\omega_v)$ for Scheme II (Fig. 2). The horizontal dashed guide line indicates a 50% probability of heterogeneous motion dominance.
The contrast c_s and spatial frequency ω_s of texture s is the same for all panels: $c_s = 0.5$ and $\omega_s = 4.9$ c/deg. (a) Subject PW; (b) subject JS.

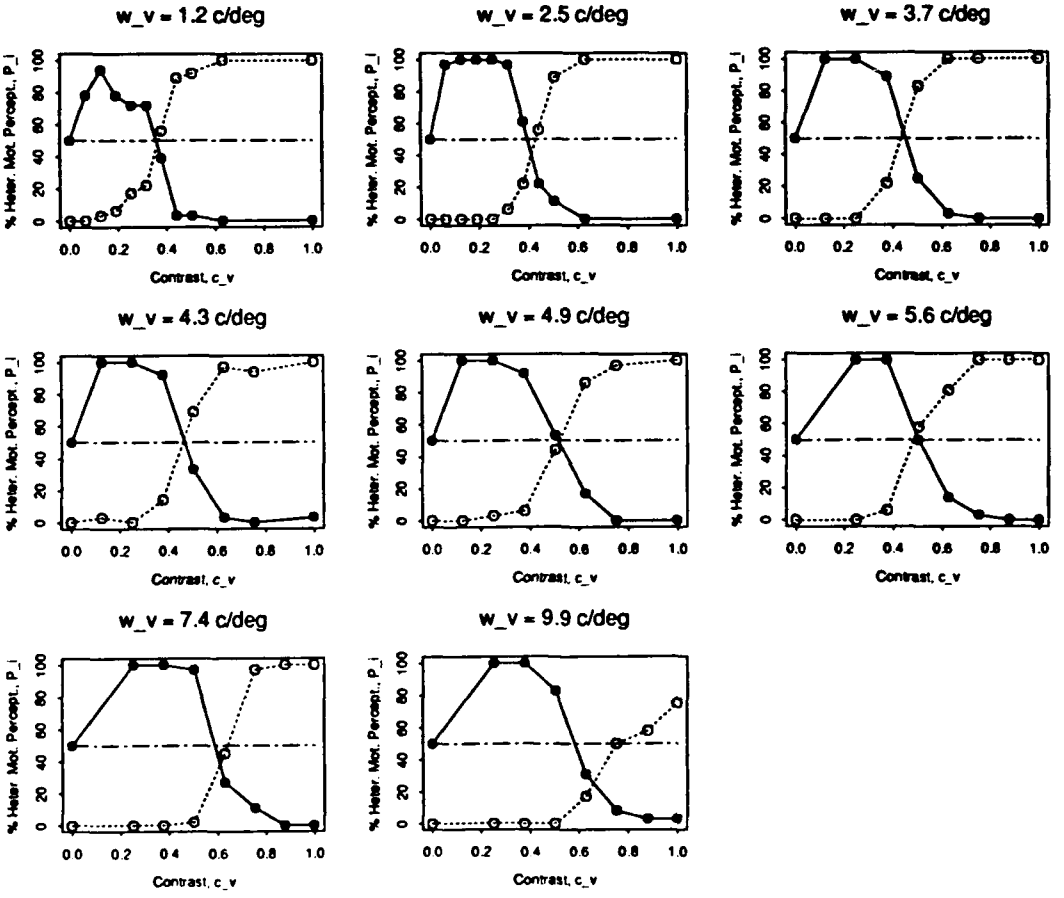


Figure 4 continued. Data for subject JS.

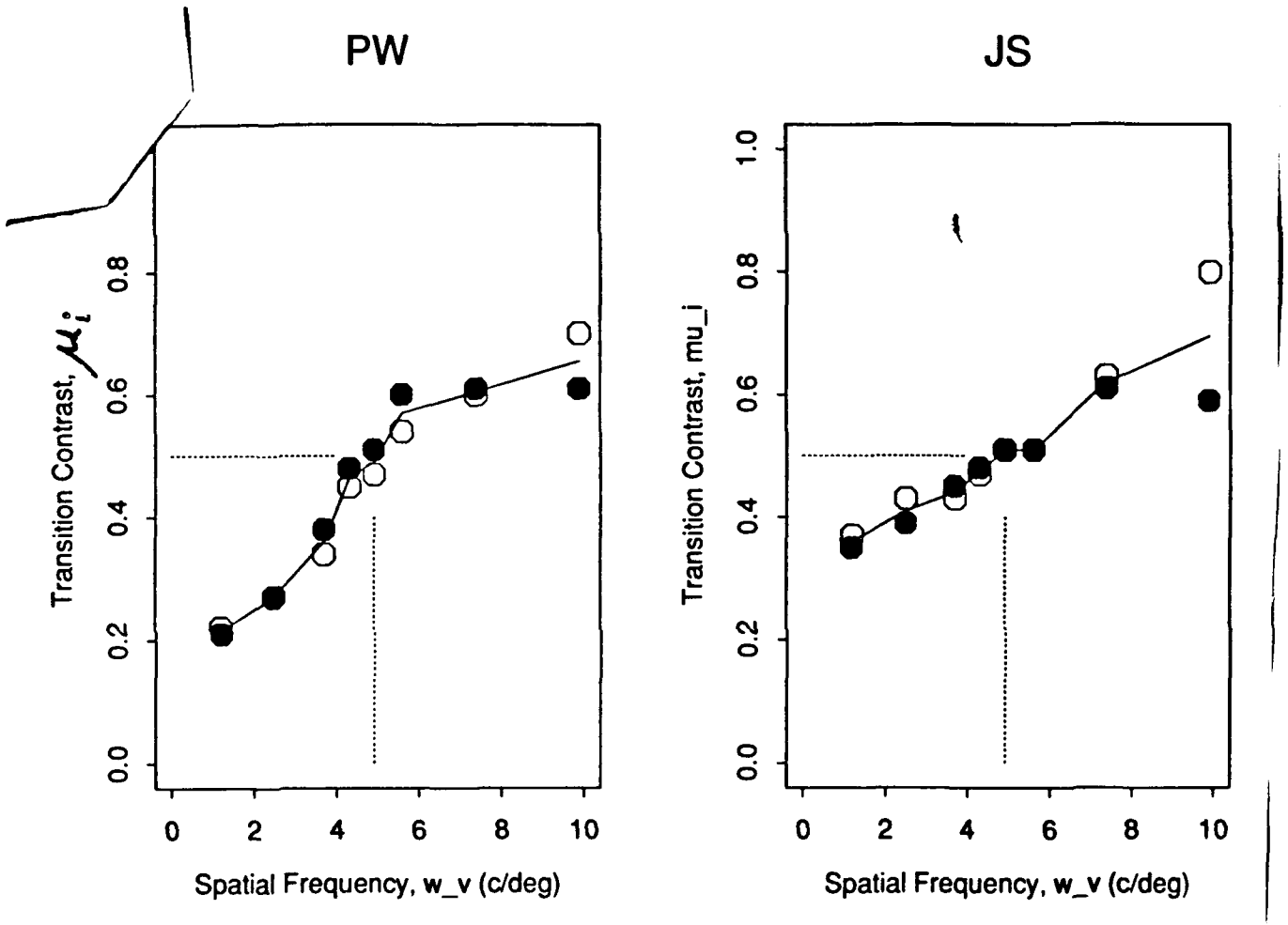


Figure 5: Transition contrasts $\mu_i(\omega_v)$ as a function of spatial frequency ω_v. Open circles for Scheme I. Filled circles for Scheme II. The vertical dashed line indicates the spatial frequency of texture s: $\omega_s = 4.9$ c/deg. The horizontal dashed guide line indicates the contrast of texture s: $c_s = 0.5$.

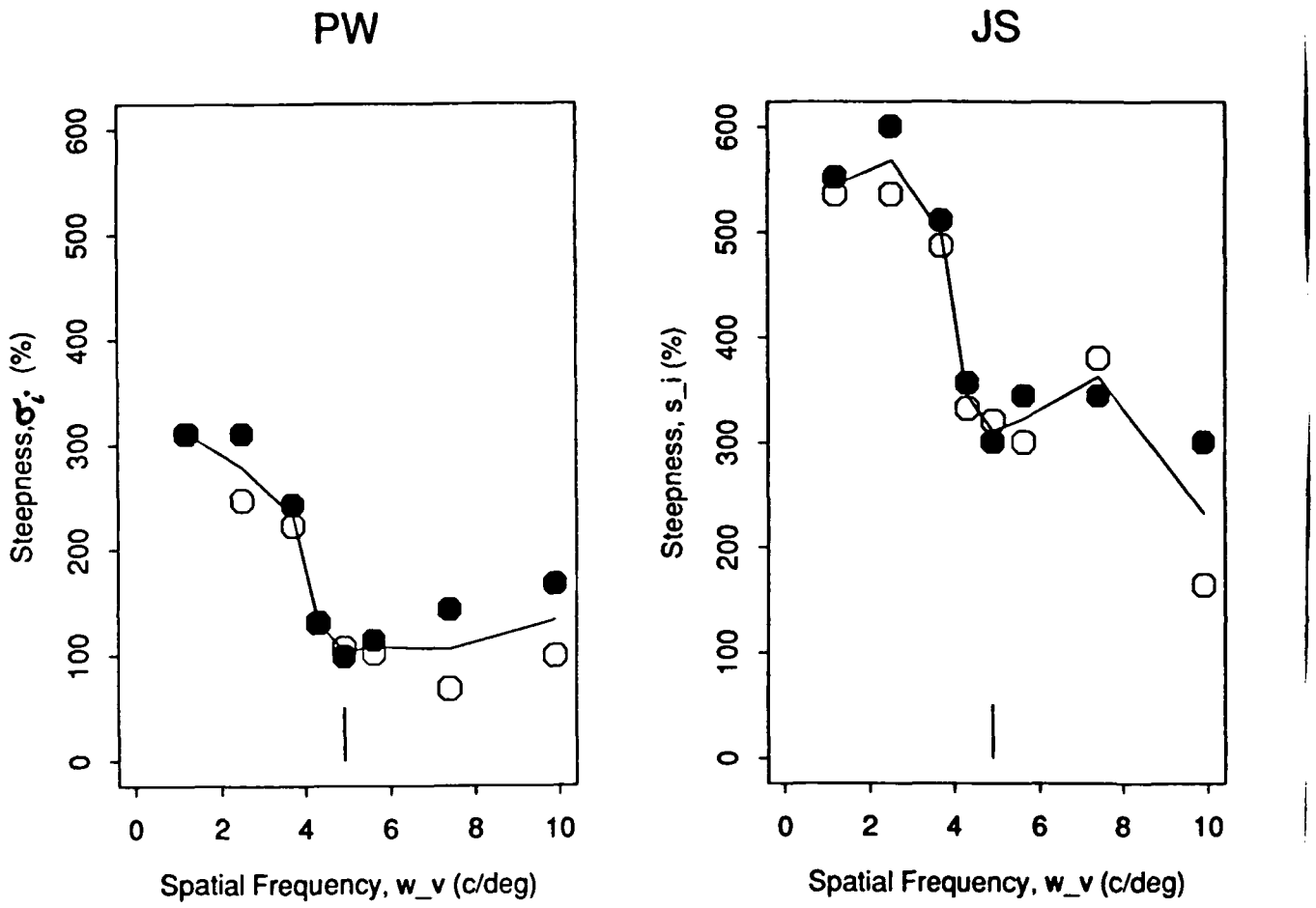


Figure 6: Steepness values $\sigma_i(\omega_v)$ as a function of spatial frequency ω_v. Open circles for Scheme I. Filled circles for Scheme II. (Note that to facilitate comparison absolute values are given!). The vertical dashed guide line indicates the spatial frequency of texture s: $\omega_s = 4.9$ c/deg.

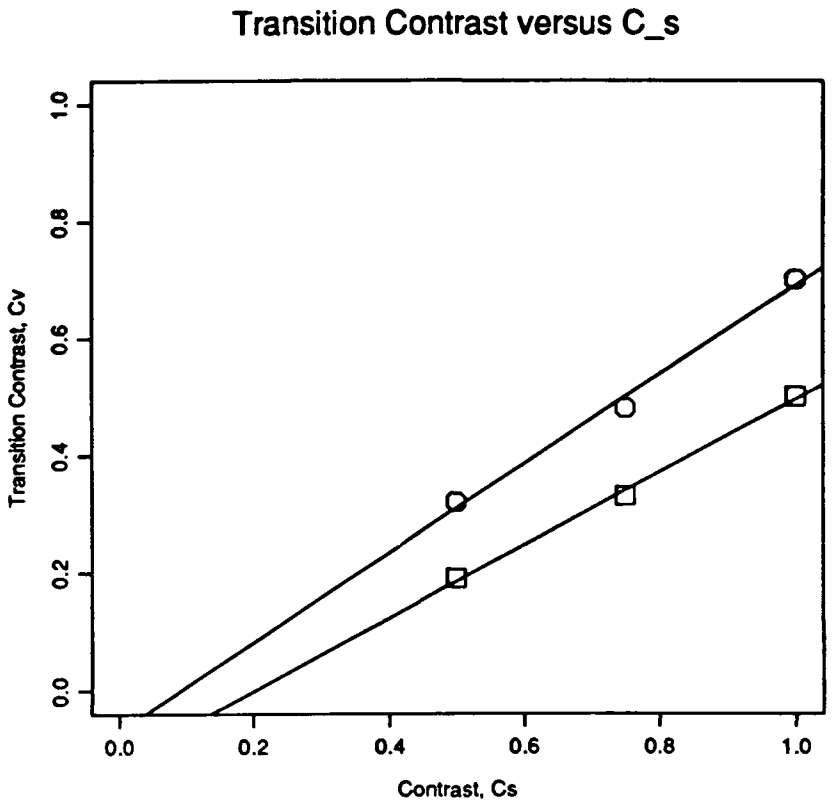


Figure 7: *The dependence of transition contrast $\mu_1(omega_v)$ on contrast c_s of texture s. The spatial frequency ω_s was 4.9 cpd, and ω_v was 1.2 cpd. Competition Scheme I was used. Circles: subject JS. Squares: subject PW. The solid lines show the best linear fit (minimizing the sum of the squared deviations).*

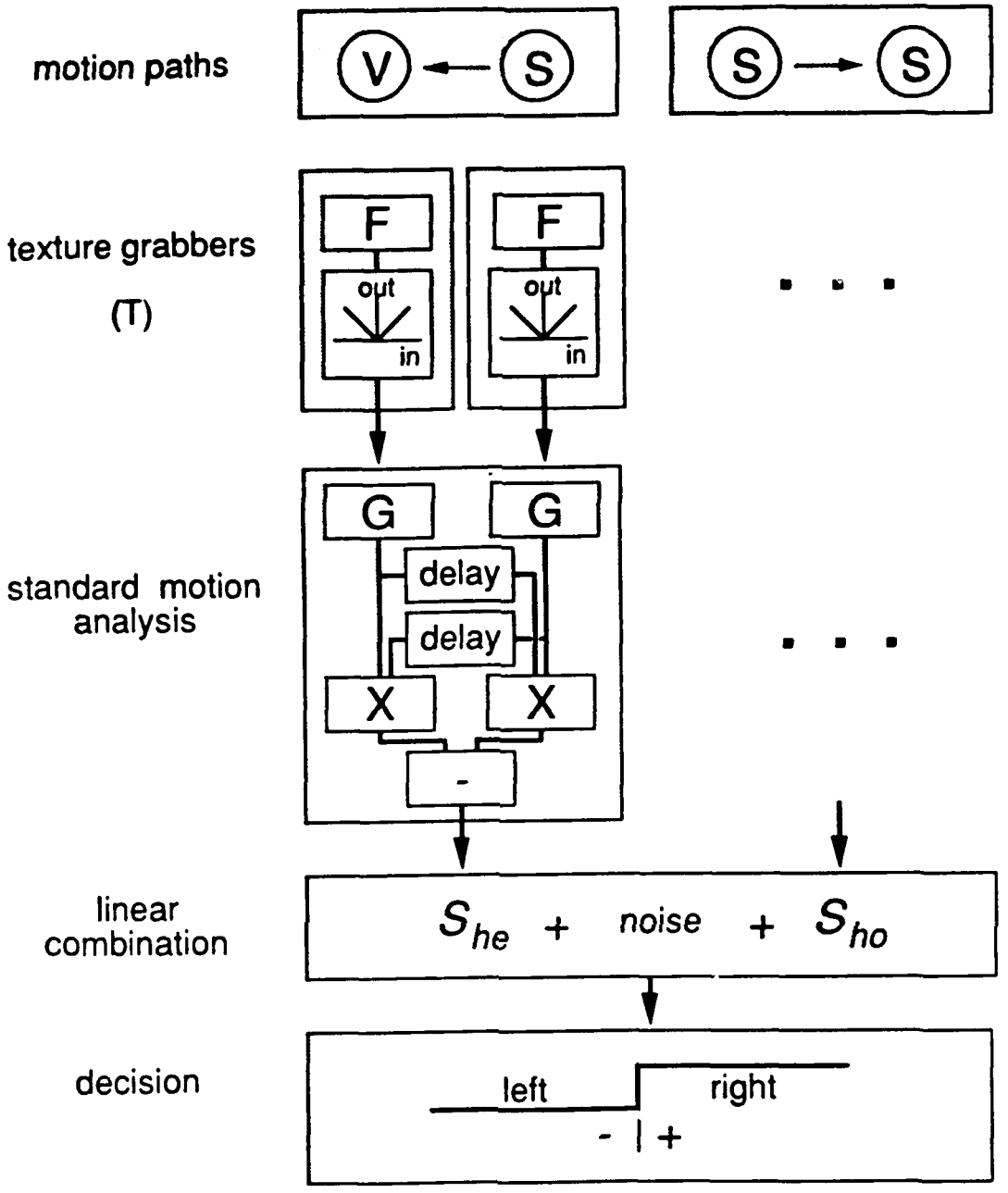


Figure 8. Diagram of a *single channel* motion computation. First stimulus contrast is extracted followed by a linear spatial filter F and rectification. The spatial filter together with the rectification is called 'texture grabber' (the first stage). The output of the texture grabber is called activity. The second stage (standard motion analysis) is basically a coincidence detector: it computes the product of the delayed activity at location 1 with the current activity at location 2. Response variability across trials is due to internal noise which is modeled by an additive noise having a standard normal density function with mean 0 and standard deviation 1. The heterogeneous path is dominant whenever the net motion strength in the direction of the heterogeneous motion path (after adding noise) is positive (decision stage).

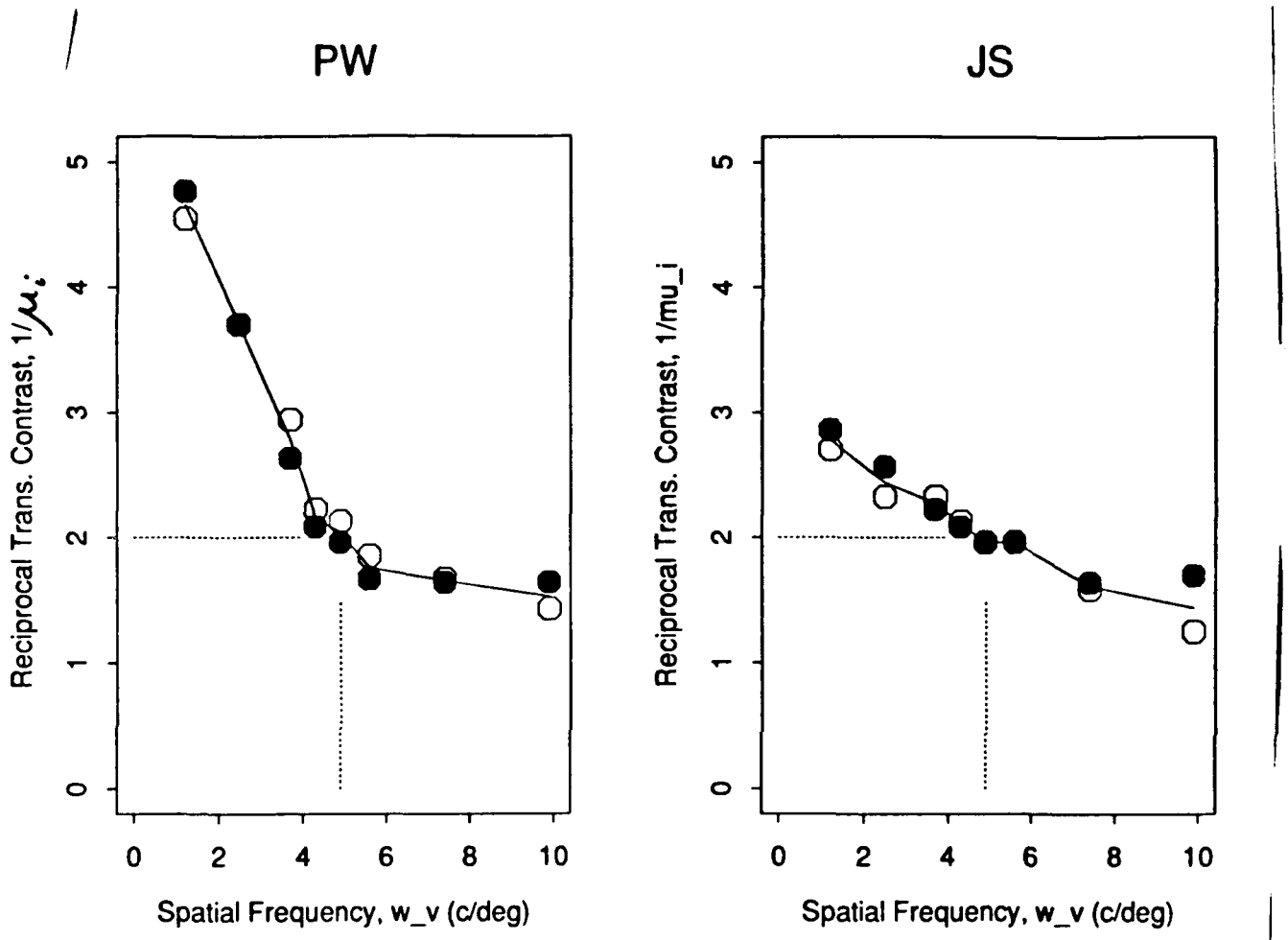


Figure 9: Reciprocal transitions $\mu_i^{-1}(\omega_v)$ as a function of spatial frequency ω_v. Open circles for Scheme I. Filled circles for Scheme II. The vertical dashed guide line indicates the spatial frequency of texture s: $\omega_s = 4.9$ c/deg. The horizontal dashed guide line indicates the contrast of texture s: $c_s = 0.5$. The solid line curve is the mean of the reciprocal transitions. In terms of the model, this curve shows the amplitude of the Fourier transform of the spatial filter $F(\omega)$ of the texture grabber involved (see Equation 13).

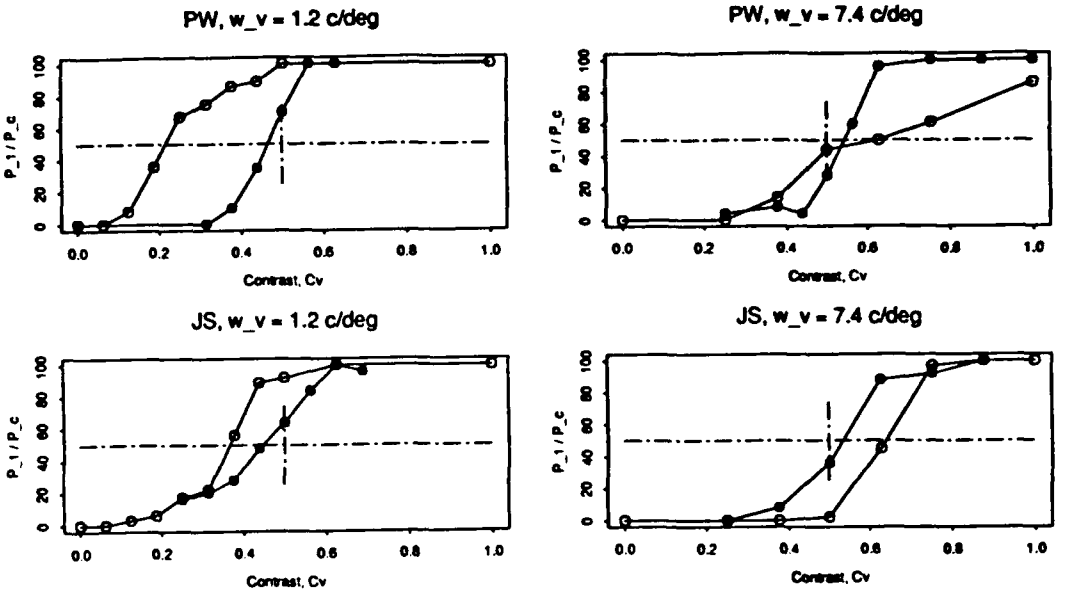


Figure 10 Results of the perceived contrast experiment. Observers compared the contrast of a grating v (spatial frequency ω_v and contrast c_v) with the contrast of texture s ($c_s = 0.5, \omega_s = 4.9$ c/deg). Shown are the probabilities P_c for judging the contrast of v higher than that of s (filled circles). The matching contrast for texture v is the crossing of the curve with the dashed 50% line.
To compare the matching contrast with the transition contrast in the motion experiment, we have shown the probabilities $P_1(c_v)$ for Scheme I (open circles).

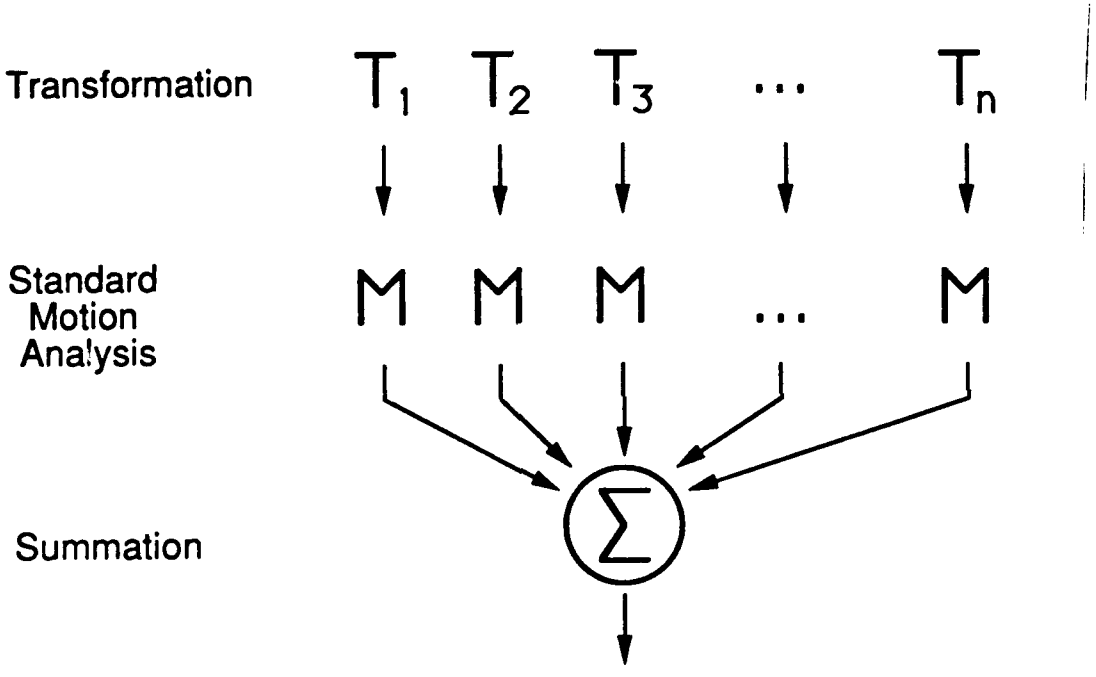


Figure 1 Multi-channel motion computation. The first stage consists of n independent transformations T_i (the texture grabbers). Transformation T_i is a non-linear transformation (*e.g.*, spatial filtering followed by rectification). The output of each transformation is called an activity representation of the optical input. Standard motion analysis (M) is applied to each of the activity representations of the input. Finally the motion strength is summed across the different channels.

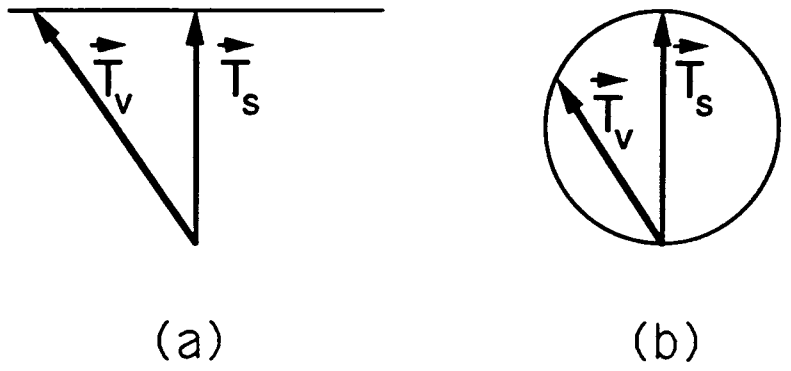


Figure 12. Solutions for transitions (path equality) in a two-dimensional T-space. Each texture in a motion path is processed by different texture grabbers. Vector $\vec{T}_v$ represents the activity of texture v in T-space, vector $\vec{T}_s$ that of s. The collection of activity vectors $\vec{T}_v$ that satisfy the constraints for path equality are given by the thin line in (a) for Scheme I and by a thin circle in (b) for Scheme II.

Visual Processing of Optic Acceleration

Peter Werkhoven[1], Herman P. Snippe[2] and Alexander Toet

Institute for Perception TNO,
Kampweg 5, 3769 DE Soesterberg,
The Netherlands

May 14, 1992

(PREPRINT — Subm. for publ. Vision Research, Feb 11 1992)

[1] Present address: New York University, Department of Psychology and Center for Neural Science, 6 Washington Place, New York, New York 10003, U.S.A.

Future address (as of May 15, 1992): Utrecht Biophysics Research Institute (UBI), Buys Ballot Laboratory, Utrecht University, Princetonplein 5, 3584 CC, Utrecht. The Netherlands.

[2] Present address: Utrecht Biophysics Research Institute (UBI), Buys Ballot Laboratory, Utrecht University, Princetonplein 5, 3584 CC, Utrecht, The Netherlands.

Abstract

We present data on the human sensitivity to optic acceleration, *i.e.*, temporal *modulations* of the *speed* and *direction* of moving objects. Modulation thresholds are measured as a function of modulation frequency and speed for different periodical *velocity vector* modulation functions using a localized target.

Evidence is presented that human detection of velocity vector modulations is *not* directly based on the acceleration signal (the temporal *derivative* of the velocity vector modulation). Instead, modulation detection is accurately described by a two stage model: a low-pass temporal filter transformation of the true velocity vector modulation followed by a *variance* detection stage.

A functional description of the first stage is a *second* order low-pass temporal filter having a characteristic time constant of 40 ms. In effect, the temporal low-pass filter is an integration of the velocity vector modulation within a temporal window of 100-140 ms. A non-trivial link of this low-pass filter stage to the temporal characteristics of standard motion detection mechanisms will be discussed.

Velocity vector modulations are detected in the second stage, whenever the variance of the filtered velocity vector exceeds a certain *threshold* variance in either the speed or direction dimension. The threshold standard deviations for this variance detection stage are estimated to be 17% for speed modulations and 9% for motion direction modulations.

Introduction

Man is capable of interacting successfully with complex dynamic environments. This ability is due primarily to powerful neural mechanisms that have evolved to process optical motion information (see Nakayama, 1985b for a survey). Therefore motion perception has been studied extensively. Psychophysical research has shown that the human visual system contains highly sensitive motion extraction mechanisms (DeBruyn and Orban, 1988; McKee, 1981; Werkhoven and Koenderink, 1991) that map spatiotemporal image structure into explicit motion information (*e.g.*, velocity and direction).

Motion perception has traditionally been studied using spatiotemporal invariant (uniform) motion stimuli. Relatively few studies have aimed at the human sensitivity to the spatiotemporal *structure* of motion fields or *velocity vector modulations.* Although previous studies helped to define methods and stimuli, none of them allowed definitive statements concerning human sensitivity to acceleration or mechanisms for detecting higher derivatives of motion (Regan *et al.*, 1986). This scarcity of studies is surprising, since in natural vision, optical motion on the retina is generally varying in both space and time even if environmental objects move at a constant speed and direction. Structured motion fields are not just an inevitable burden for our visual system. In fact, it has been shown that the *spatial* structure (Koenderink, 1986) and *temporal* structure (Arnspang, 1988) of optical motion fields are of major importance to the visual agent and are closely related to egomotion and 3D shape extraction.

The study presented here focuses on the human sensitivity to temporal velocity vector modulations, that is, the ability to detect temporal variations in speed or direction (called *optic acceleration*).

The Paradigm

A fundamental and intriguing question to be answered is: Does the human visual system contain specific acceleration detectors? In other words, do human observers *directly* assess the optic acceleration of a moving object (the temporal *derivative* of the velocity vector function) or do they *indirectly* infer optic acceleration from variations in the perceived velocity along its trajectory (by sampling velocities at different times)?

This question strongly resembles a classic debate in the study of uniform motion perception: Are human observers able to *directly* sense optical motion, or do they infer motion *indirectly* from the variance in object position over time? Nakayama and Tyler (1981) have answered the latter question using a target with a periodically (sinusoidally) modulated position in time. They measured modulation threshold amplitudes as a function of the frequency (inverse period) of the position modulation function. They argued that modulation threshold amplitudes would be independent of the modulation frequency when motion was inferred from the variance in position. However, when motion was assessed directly (*e.g.*, the temporal derivative of the

position modulation function) threshold modulation amplitudes were expected to decrease with increasing modulation frequency.

For low modulation frequencies (< 2 Hz), Nakayama and Tyler found strong experimental support for a direct assessment of motion. Modulation thresholds did not show an invariance when expressed in terms of displacement. For higher frequencies (> 2 Hz), Nakayama and Tyler found deviations from the expected dependence of modulation thresholds on frequency, presumably as a consequence of some finite temporal integration of the motion signal in the human motion system.

To examine optic acceleration, we adopt this elegant paradigm used by Nakayama and Tyler substituting velocity modulations for position modulations. That is, we use a target with a velocity vector modulated in time around a certain mean velocity vector and measure threshold amplitudes for the detection of velocity vector modulations as a function of the modulation frequency. We study velocity vector modulations both in the direction of the velocity vector (speed modulation) and orthogonal to the velocity vector (direction modulation). Invariant modulation thresholds as a function of modulation frequency would indicate an *indirect* detection of motion modulation or optic acceleration.

General Stimulus Considerations

The choice of an adequate stimulus to be used in a study on motion modulation detection is not trivial. It is important to design the modulation detection experiment such that detection cannot take place *outside* the motion system in other dimensions than speed or direction. In the following we list a few considerations regarding some widely used stimuli in motion experiments.

Sine Wave Gratings

Sinewave luminance gratings are a powerful tool for studying linear systems and also for studying motion perception. However, the use of moving sine wave gratings leads to several problems. First, local speed and local temporal frequency are inherently confounded. As a result, a speed modulation of a moving sine wave grating might be detected outside the motion system as a local modulation of stimulus temporal frequency. For example, a detector with a spatiotemporal *separable* response function, thus not tuned to speed at all, would be sufficient. Second, with a one-dimensional spatial pattern, such as a sine wave grating, it is not at all obvious how one could study motion *direction* modulations. Third, moving sine wave gratings allow extensive spatial integration by the motion detection system. This property makes all spatially extended moving patterns especially unattractive to study spatially *local* modulations in speed or direction. Fourth, spatially extended moving patterns inherently stimulate motion detectors at a range of eccentricities. Thus, a study of motion sensitivity as a function of eccentricity cannot be specific.

Random Pixel Arrays

Another visual stimulus often used in studies on motion perception is a random pixel array or 'Julesz pattern' (Julesz, 1971). An important property of a Julesz pattern is that its power spectrum is flat. Therefore, a moving Julesz pattern with a modulated speed function would not yield the temporal frequency cue discussed above. However, human sensitivity to temporal modulations is limited by the flicker fusion frequency. As a result of this cut-off frequency for temporal modulations, the sensed energy of a moving Julesz pattern decreases when speed increases. That is, when speed increases, an increasing proportion of the spectral components of the moving pattern would yield temporal frequencies beyond the fusion frequency, thus conceivably reducing the apparent contrast of the stimulus. Thus, speed modulation for Julesz patterns may provide the observer with an apparent contrast modulation as a cue.

Furthermore, Julesz patterns are spatially extended. Hence, they yield similar problems for the study of motion modulations as discussed above for sine wave gratings.

Localized Targets

We have discussed a few extraneous cues associated with spatially extended stimuli. Many of these problems are circumvented when using strongly localized targets, such as dots. A moving dot allows for the study of *local* motion perception (restricted spatial integration) and for the control of eccentricity of presentation. Furthermore, local temporal frequency modulation is not a cue for motion modulation detection.

However, an increase in dot speed can yield an increase in apparent spatial stimulus extent (if the visual system integrates the stimulus over a fixed window in time), and also a decrease in apparent contrast (if the visual system integrates the stimulus over a fixed window in space). To get some grip on the possible contributions of these extraneous cues, we studied motion modulation sensitivity using moving dot targets and blob targets (spatially blurred dots).

Methods

Method Speed Modulations

This section describes the method for our study on the human sensitivity to temporal modulations of motion *speed.*

Stimulus Specifications

The stimulus consisted of a moving luminous dot (well above detection threshold) of 1 mm diameter. The dot projected on the screen of a CRT was blurred by a sheet of diffusing material which was placed directly in front of the CRT screen. We estimate the standard deviation of the resulting isotropic luminance 'blob' at 1.5 cm, thus its full width at half maximum (FWHM) at about 3-4 cm. The dot moved horizontally across the screen at a variable (modulated) speed from the leftmost point to the rightmost point of a horizontal trajectory across a distance

d_0. This single left-to-right motion is called a *sweep* (the distance d_0 is the sweep-length). When it reached the right end on its trajectory, the dot returned to the far left position on the trajectory and continued its motion (the next sweep). The time to finish one sweep is called the sweep-time (t_0). At a viewing distance d_v, the average dot speed v_0 was:

$$v_0 = t_0^{-1} \arctan(d_0/d_v). \tag{1}$$

One motion stimulus presentation consisted of 4 sweeps. Thus, the presentation time was $4t_0$. The dot speed was modulated in time yielding a non-uniform periodic speed function $v_x(t)$ with modulation frequency ω. Speed modulation functions $v_x(t)$ were either periodic (symmetric) triangular functions $\Lambda(t)$ or periodic block functions $\Pi(t)$ (see Fig. 1). The amplitude dv_x of the modulation functions was varied but was always smaller than the average speed v_0, such that the dot speed was always positive. The phase of the periodic modulation function at the start of the stimulus presentation was randomized.

— Figure 1 about here —

In addition to the moving dot, we also provided the observer with a stationary fixation dot (a green LED), placed at a distance equal to the sweep-length d_0 above the center of the horizontal trajectory, thus making eccentricity of presentation (ϵ) about equal to the length of the trajectory of the moving dot. The sweep-time for a particular experiment was taken to be such that one sweep contained a few cycles of the speed modulation. Hence, for low temporal modulation frequencies examined, a longer sweep-time was required. The parameters as set in the different modulation experiments are specified in separate parameter tables in the Result section.

In the main experiments, speed covaried with the eccentricity of the moving dot. To study the effects of eccentricity and speed independently, we ran two control experiments. In one, we varied eccentricity but kept the viewing distance constant. In the second, we varied speed but kept eccentricity constant.

Apparatus

The speed modulation functions were generated by manipulating the position of the beam of a HP 1321A high speed graphic display (P31 phosphor).

The beam produced a 1 mm diameter luminous dot on the screen (well above detection threshold). A Wavetek 185, 5 MHz, function generator produced a saw-tooth horizontal position signal $x(t)$, as a function of time t, which was fed into the X-channel of the HP 1321A. The horizontal position of the dot was linear with this signal. Hence, the dot moved from left to right across the screen until the saw-tooth reached its maximum (finishing one *sweep*), at which point it returned (invisible) to the far left and started to traverse the screen again (the next sweep). The amplitude of $x(t)$ (and thus the sweep-length) across the screen was

constant. For a constant period λ of this saw-tooth signal in time, the dot crossed the screen at a constant speed, determined by the temporal derivative of x(t), and thus proportional with the reciprocal period $1/\lambda$ of x(t). The reciprocal period $1/\lambda$ of $x(t)$ (and thus the dot speed) was modulated in time by a periodic modulation function $v_x(t)$ with (modulation) frequency ω (using a HP 3325A synthesizer function generator). The modulation function $v_x(t)$ was either a triangular function $\Lambda(t)$ or a block function $\Pi(t)$ (see Fig. 1). Speed modulation $v_x(t)$ varied around an average speed v_0 with an amplitude dv_x (see Fig. 1). This set-up allowed easy adjustment of the average speed, amplitude and frequency of the speed modulation function.

The Importance of Visual Fixation

Pilot experiments showed that visual fixation during modulation detection experiments is critical. Observers reported to have *no difficulties* in detecting modulations when tracking the moving dot for conditions where detection *failed* under visual fixation. Obviously, pursuit eye movements facilitate modulation detection. It is well-known that the pursuit system is quite slow (cut-off frequency at about 1 Hz). For speed modulation frequencies higher than 1 Hz, observers could not follow the exact speed modulation, but might track the dot at its average speed. The actual speed modulation would then become apparent as a displacement in the retinal coordinate frame. Thus, allowing the observers to track the dot would provide them with a displacement cue, resulting in modulation detection outside the motion system. In order to eliminate this cue, visual fixation is crucial.

Note that in much of the older literature (*e.g.*, Hick, 1950), but also in more recent literature (Burr *et al.*, 1986) no mention of the observers fixation condition is made.

Procedural Information

Speed modulation thresholds were measured in a modulation *detection* experiment. In one session, observers viewed 18 stimulus presentations of a modulated speed function $v(t)$ (with an average speed v_0 and a modulation amplitude dv_x) and 18 presentations of a unmodulated (uniform) speed function (having a constant speed v_0). The order of presentation for these 36 trials in a session was randomized, as was the phase of the modulation function for the trials that contained the speed modulation. The task of the observers was to indicate for each stimulus presentation whether they perceived a modulated or an unmodulated motion in time.

Usually 4-5 sessions with different adequately chosen modulation amplitudes were sufficient to determine the speed modulation detection threshold by data interpolation. We defined the speed modulation threshold W_v as the relative modulation amplitude dv_x/v_0 at threshold performance (yielding 80% correct answers). Measurements were performed binocularly with natural pupils in a darkened room. No feedback was provided in either experiment.

In one of the control experiments we studied speed *discrimination* using the present experimental set-up. In a session for speed discrimination , observers viewed uniform speed functions with a constant speed that was either higher ($v_0 + dv_x$) or lower ($v_0 - dv_x$) than the average speed v_0 of the ensemble of presentations. Observers indicated whether the perceived speed was high or low. Before a session started, the motion stimuli were shown on request to build

an internal representation of the high and low speeds. The procedure for determining speed discrimination thresholds was otherwise similar to the procedure for the modulation detection experiment.

Subjects

Five subjects with normal or corrected-to-normal vision participated in the experiments. Three subjects, HS, PW, and AT are authors of this paper and had foreknowledge of the design, and are experienced observers in psychophysical experiments involving optic motion. The results of these main subjects are presented. The general findings were confirmed by two naive subjects, working on an hourly fee. There was no obvious correlation between subject experience and threshold values.

Method Direction Modulations

This section describes the method for experiments on *direction* modulation detection.

Stimulus

Similar to speed modulation functions, the direction modulation functions were generated by manipulating the position of the beam of the HP 1321A high speed graphic display. However, for direction modulation functions, both the horizontal and vertical position of the dot were manipulated.

The time dependent horizontal position $x(t)$ of the dot (the X-channel of the HP 1321A) was driven by a HP 3325A synthesiser function generator. This generator produced a saw-tooth signal x(t) with a period λ that determined the sweeptime t_0 and an amplitude that determined the sweep-length d_0. For this direction modulation experiment, the period and amplitude of $x(t)$ were constant during a stimulus presentation, resulting in a constant horizontal speed $v_x(t) = v_0 = d_0/t_0$.

The time dependent vertical position $y(t)$ of the dot (the Y-channel of the HP 1321A) was driven by Wavetek 185 function generator. The $y(t)$ signal determined the direction modulation. The position functions $y(t)$ were periodic with frequency ω. The vertical speed function $v_y(t)$ was simply the temporal derivative of vertical position $y(t)$. Thus, the modulation frequency was ω. The average vertical speed was zero. The amplitude of the vertical speed modulation function $v_y(t)$ is written as dv_y. As a result, the speed of motion $v(t)$ of the dot was:

$$v(t) = v_0\sqrt{1 + \frac{v_y^2(t)}{v_0^2}}. \tag{2}$$

For a small vertical speed $v_y(t)$ relative to the horizontal speed v_0 ($v_y(t) \ll v_0$), the average speed was approximately constant ($v(t) \approx v_0$). The direction $\theta(t)$ of motion as a function of time t is approximately linear in $v_y(t)$ when $v_y(t) \ll v_0$:

$$\theta(t) = \arctan(\frac{v_y(t)}{v_0}) \approx \frac{v_y(t)}{v_0}. \quad (3)$$

The average motion direction θ_0 in all experiments was horizontal: $\theta_0 = 0$. The amplitude of the direction modulation function $\theta(t)$ is $d\theta = \arctan(dv_y/v_0)$.

Triangular position functions $y(t) = \Lambda(t)$ resulted in block shaped direction modulation functions $\theta(t) = \Pi(t)$. Sinusoidal functions $y(t) = \Omega(t)$ resulted in sinusoidal direction modulations $\theta(t) = \Omega(t)$ (integrated $y(t)$ functions), but shifted a phase $\pi/2$ backwards in time. Finally, block wave position functions $y(t) = \Pi(t)$ resulted in pulse shaped direction modulation functions $\theta(t) = \delta(t)$. An illustration of these position modulation functions and resulting direction modulation functions is shown in Fig. 1.

Procedure

The procedure was identical to the procedure for speed modulation detection experiments. Observers indicated for each motion stimulus whether the motion was modulated (non uniform) or not. Two main observers participated in the direction modulation experiments (HS and AT). Two naive subjects confirmed the findings for the main subjects.

Modulation direction thresholds are defined as the direction modulation amplitude $d\theta$ yielding threshold performance (80% correct answers).

Speed Modulation Detection

Speed Modulation Detection Dependence on Modulation Frequency

Results

Speed modulation detection thresholds W_v as a function of modulation frequency ω for two different speed modulation functions and different speeds v_0 are presented in Fig. 2.

— Figure 2 about here —

The parameter settings for different average dot speeds v_0 are listed in Table 1.

— Table 1 about here —

Since the data were very similar for the three main observers (PW, HS and AT), we averaged modulation detection thresholds W_v for this presentation. The modulation thresholds of Fig. 2 are presented as (relative) speed modulation thresholds (speed modulation Weber fractions $W_v = dv_x/v_0$). Triangular symbols indicate the triangular speed modulation function $\Lambda(t)$ and square symbols indicate the block modulation function $\Pi(t)$. Open symbols indicate that the moving target was blob like. Closed symbols indicate dot targets.

Consider triangular modulation functions. For low modulation frequencies ($\omega \leq 2$ Hz) speed modulation thresholds for very different conditions (1 deg/s dot targets at 0.25, 0.5 and 1 Hz, 1.7 deg/s blob targets at 1 and 2 Hz, and 15 deg/s blob targets at 2 Hz) are identical within measurement error (approximately 32%). This suggests that speed modulation detection thresholds at low frequencies are *constant* and *independent* of frequency, speed and target shape. However, speed modulation detection threshold values do depend on the shape of the modulation function used. Thresholds for the triangular speed modulation functions $\Lambda(t)$ are approximately a factor 1.8 higher (averaged over speeds, modulation frequencies and subjects) than the thresholds (17%) for the block modulation function $\Pi(t)$.

At high modulation frequencies ($\omega > 2$ Hz) the speed modulation thresholds in Fig. 2 rise strongly with increasing modulation frequency for both triangular and block shaped modulation functions.

Discussion: Threshold Invariance at Low Modulation Frequencies

The frequency independence of modulation thresholds for low modulation frequencies strongly supports the hypothesis that modulation detection is based on the *magnitude* of the speed modulation signal. The modulation magnitude is *independent* of modulation frequency. A detection mechanism based on the difference in maximum and minimum speeds of the speed modulation function is indeed expected to yield constant thresholds, independent of modulation frequency.

The invariance of thresholds for low frequencies rules out the hypothesis that speed modulation detection is based on the magnitude of the optic *acceleration* signal. The optic acceleration signal is the *temporal derivative* of the speed modulation signal. Hence, its magnitude is linear with the modulation frequency. Therefore, a detection based on the acceleration magnitude is expected to improve with increasing modulation frequency. A hypothetical acceleration detector (requiring a constant acceleration threshold for detection) would yield a hyperbolic (inverse linear) *decrease* of speed modulation threshold in Fig. 2, which is not supported by the data.

The low frequency plateau in Fig. 2 rules out another hypothesis saying that observers base detection on the *spatial excursions* of the moving dot with respect to its average path (*i.e.*, the path of constant speed v_0). According to this hypothesis, the speed modulations are detected whenever the excursions exceed a certain excursion threshold. The magnitude of the spatial excursion is the *temporal integral* of the speed signal and is linear with speed and with the period of temporal modulation. Thus, the 'excursion' hypothesis predicts that speed modulation thresholds decrease with *decreasing* modulation frequency. This prediction is inconsistent with the finding that thresholds are constant for low modulation frequencies (see Fig. 2).

In conclusion, the threshold invariance at low modulation frequencies strongly support the view that human speed modulation detection is based on the speed signal itself (the relative *magnitude* dv_x/v_0 of the speed modulation function $v(t)$), and not on the temporal integral of $v(t)$ (position), or the temporal derivative of $v(t)$ (acceleration).

Discussion: Low-pass Temporal Filtering at High Frequencies

At high modulation frequencies ($\omega > 2$ Hz) the speed modulation thresholds in Fig. 2 rise strongly with increasing modulation frequency. Apparently, the magnitude of the modulation function is reduced at high frequencies. Phenomenologically this can be understood by assuming that human speed modulation detection is based on a temporally blurred (low pass filtered) version of the true (physical) speed function. A temporal low-pass filtering of the modulation function reduces the energy of high frequency modulation functions yielding increased thresholds amplitudes.

In the Model Section, we elaborate on a two-stage modulation detection model. The first stage consists of a nth order low-pass temporal filter operating on the speed modulation function. The second stage is decision stage based on the filtered modulation signal of stage one. The parameters that specify the first (temporal filter) stage and the second (decision) stage are derived based on data presented here and in the following sections.

Discussion: Wave Forms

At a given modulation frequency, thresholds for triangular speed modulation functions are a factor 1.8 higher than the thresholds for the block modulation functions. At high modulation frequencies, this finding can be understood by considering the fundamental low-pass characteristic of the modulation detection threshold functions obtained. We estimate that for frequencies $\omega > 2$ Hz only the fundamental frequency is passed through, even at 100% modulation depth (note that because of the symmetry of the modulation functions used, only the odd harmonics are present). This ratio of 1.8 for the relative thresholds of triangular and block modulation functions is in reasonable agreement with the ratio $\pi/2 \approx 1.6$ of the amplitudes of the fundamental frequencies of the two modulation functions. The fact that the fundamental frequency component dominates the percept is supported by informal and introspective reports of our observers. They reported not to be able to discriminate the temporal *pattern* of block and triangular modulation functions with equal apparent modulation depth at frequencies higher than 2 Hz, whereas they were able to perform such a discrimination for the 1 Hz modulation. A further study of this issue may be of interest.

At low modulation frequencies (*e.g.*, 1 Hz) however, the dependence of thresholds on the form of the modulation function *cannot* be understood by considering low-pass temporal filtering, because at these modulation frequencies this filter is fast enough to follow the physical speed function. For low modulation frequencies, the filtered speed modulation signal will pass through the temporal filter unaffected for both triangular and block modulation functions. This observation reveals information about the type of detection that operates on the low-pass filtered modulation function. It strongly suggests that the modulation detection cannot be

based simply on the amplitude (or peak) of the modulation function (yielding equal thresholds for triangular and block modulation functions).

In the Model Section we show that this apparent discrepancy can be resolved by (1) taking into account the effects of probability summation on peak detection (note that a block modulation function spends much more time at large, near threshold, excursions than a triangular modulation function), or (2) by assuming a *variance* detection of the filtered speed signal instead of a peak detection. Note that a block wave modulation function has a variance three times that of a triangular wave of the same amplitude. Hence, the amplitude (modulation depth) of the triangular function has to be $\sqrt{3} \approx 1.7$ times that of the block function for threshold performance in a variance detector.

Discussion: Blob Targets versus Dot Targets

We argued in the introduction that speed modulations are confounded with changes in the spatiotemporal power spectrum yielding an apparent contrast cue for speed modulation detection. Also, the perceived spatial structure might change with speed. For example, a dot at constant speed might be perceived as a horizontal bar for high frequency speed modulations. However, the correspondence of speed modulation thresholds for luminous dots and blurred blobs (see Fig. 2) (at 1 Hz for the triangular modulation function) shows that the exact spatial structure (or frequency spectrum) of the stimuli is not critical for the value of the low-frequency speed modulation Weber-fraction.

Speed Modulation Detection v. Speed Discrimination

Motivation

The threshold amplitudes (≈17%) for the block modulation functions (open squares in Fig. 2) are in excellent agreement with recent data reported by Snowden and Braddick (1991) obtained for speed modulated random dot patterns.

Perhaps surprisingly, however, thresholds for speed *modulation detection* are much higher (typically a factor 3-4) than thresholds found in speed *discrimination* experiments (DeBruyn and Orban, 1988; McKee, 1981). Unfortunately, speed discrimination thresholds are often dependent on the specific experimental conditions. In order to compare modulation detection with discrimination thresholds, we performed a speed discrimination experiment using the same experimental set-up as for modulation detection. Observers had to indicate whether the perceived speed of a uniformly moving dot was high ($v_0 + dv_x$) or low ($v_0 + dv_x$). As for modulation detection thresholds, speed discrimination thresholds are expressed relative to the average speed: $W_v = dv_x/v_0$.

Results

We measured two speed discrimination thresholds at different presentation times of the motion stimulus. Each uniform motion stimulus in our two-interval discrimination experiment

was shown for an interval duration $T = 125$ ms or $T = 1000$ ms. The corresponding thresholds are shown in Fig. 2 ('+' symbols) at their 'equivalent' temporal frequencies $\omega = 1/(2T)$. This facilitates a comparison of discrimination thresholds with modulation detection thresholds at frequency ω for which each speed interval of the modulation function was shown for $1/(2\omega)$ s.

Results (see '+' symbols in Fig. 2) show that speed discrimination thresholds (6%) are indeed much lower than speed modulation detection thresholds (17% for block wave modulation functions). The 6% speed discrimination thresholds were independent of presentation time.

Discussion

At the longest presentation time each uniform speed was shown 1000 ms in the speed discrimination experiment, yielding a 6% threshold. It is interesting to compare this 6% speed discrimination threshold with the 17% speed modulation threshold for block modulation functions at 0.5 Hz modulation frequency. For a block shaped modulation function at this frequency, the presentation time of each speed interval of the block function was also 1000 ms. Thus, although the different speeds in both experiments were presented at equal (long) time intervals, the thresholds are markedly different.

The high thresholds for modulation detection may be a consequence of a fundamental problem observers have in *segmenting* the modulated motion stimulus into high and low speed intervals when speed itself is the only segmentation cue, as originally proposed by Snowden and Braddick (1991). However, we propose an alternative explanation (as discussed in detail in the Model and General Discussion sections): High thresholds for modulation detection may be caused by the uncertainty (of observers) about the phase of the speed modulation function.

The Cut-off Frequency Dependence on Speed

Motivation

The cut-off frequency $\omega_c(v_0)$ is defined as the modulation frequency yielding threshold detection performance (80% correct answers) for a given average speed v_0 and modulation amplitude dv_x. The data in Figure 2 suggest that this cut-off frequency is a function of speed. For example, for the lowest speed tested (1.7 deg/s), modulation detection thresholds increase somewhat faster for increasing modulation frequency (*e.g.*, at 4 Hz) than the threshold for an average speed of 15 deg/s.

To study this issue further, we measured cut-off frequencies for a wide range of average speeds v_0. To facilitate a comparison of our data with the cut-off frequencies for random dot patterns used in the experiment of Snowden and Braddick (1991), we removed the diffusing screen yielding a luminous dot as a target. The spatial power spectra of dot targets and random dot patterns are comparable.

Results

We measured cut-off frequencies $\omega_c(v_0)$ for a wide range of average speeds v_0 by measuring

the percentage of correct responses as a function of modulation frequency at modulation depth $dv_x = 100\%$ for the triangular speed modulation function $\Lambda(t)$. The speeds and corresponding parameter settings are listed in Table 2. It should be noted that target speed was varied by varying the viewing distance d_v (although for two conditions the sweep-length d_0 was slightly adjusted).

— Table 2 about here —

The closed symbols in Figure 3 are cut-off frequencies for dot targets and show a clear increase of cut-off frequency with stimulus speed.

— Figure 3 about here —

Open symbols are cut-off frequencies for blob targets and are extrapolated from the thresholds for triangular modulation functions in Fig. 2 using a temporal low-pass filter that is justified and specified in the Model Section. Because these extrapolated data for observer AT and HS were very similar, we averaged them for this presentation.

Discussion

We fitted the dependence of the cut-off frequency $\omega_c(v_0)$ for dot targets (closed symbols) on speed v_0 to a power function:

$$\omega_c(v_0) \propto v_0^{\alpha}. \tag{4}$$

and estimated the power exponent $\alpha = 0.3 - 0.35$. In the General Discussion section, we discuss this power law in terms of well known properties of elementary motion detectors.

A comparison of the cut-off frequencies for blob and dot targets shows that only for the highest speeds used ($v_0 > 7.5$ deg/s) the cut-off frequency becomes pattern-dependent (see also Watson *et al.*, 1986). This is consistent with introspective reports saying that, for still higher average dot speeds of the modulation, the percept was a 'string of beads'. The 'beads' presumably correspond to the places where the stimulus comes to an instantaneous standstill, thus allowing a significant luminance build-up over time in a small spatial region.

Disentangling Viewing Distance and Eccentricity

Motivation

In this section we report on a control experiment to test our claim that the high frequency cut-off we find for our speed modulation thresholds (see Fig. 3) is caused by low-pass *temporal*

filtering. In Fig. 3 we showed that the temporal cut-off frequency depends only weakly on stimulus speed (4-8 Hz for speeds less than 7 deg/s). Therefore, it is tempting to assume a temporal frequency limit for the speed modulation detection system. However, to support this conclusion we have to tackle the following problem.

In the above experiment, speed was varied by varying the viewing distance (see Table 2), thus covarying the eccentricity of presentation with stimulus speed. The spatial grain size of the visual system increases approximately linear with increasing eccentricity (Watson, 1987), such that the spatial resolution for the spatial speed variations of our stimuli decreases as a function of eccentricity. Therefore, one could claim that the (near) invariance of the temporal frequency cut-off can also be explained by a constant spatial frequency limit with respect to the grain size of the visual system at the eccentricity of presentation of the motion stimulus.

To disentangle the effects of viewing distance and eccentricity we measured the cut-off frequency at a fixed viewing distance and stimulus speed, but at different eccentricities.

Results

We measured the cut-off frequency at a fixed viewing distance (d_v = 2.40 m) and stimulus speed (4 deg/s) for different eccentricities of presentation (ϵ = 0.5, 5, 10 and 15 deg). The sweep-length d_0 is 42 cm and the sweep-time t_0 is 2.5 s. Cut-off frequencies for a triangular speed modulation function with 80% modulation depth are presented in Table 3.

— *Table 3 about here* —

Table 3 shows that the cut-off frequencies are virtually identical at all eccentricities.

Discussion

A correct explanation for the approximately invariant cut-off frequencies is indeed in terms of a temporal high-frequency cut-off, and not in terms of an eccentricity-scaled spatial resolution limit. Of course this temporal frequency limit *can* be described as a spatial limit in units that scale with stimulus *speed*. However, because of the scaling in human motion vision of the spatial grain size with stimulus speed, we believe that such a description is equivalent to our explanation in terms of a temporal resolution limit.

Disentangling Viewing Distance and Speed

Motivation

The cut-off frequencies in Table 3 at constant speed but varying eccentricity are invariant, whereas the cut-off frequencies in Fig. 3 at covarying stimulus speed and eccentricity do show a slight (though systematic) variation. We hypothesized that this small variation in cut-off

frequencies depends on the stimulus speed. We tested this hypothesis explicitly by measuring cut-off frequencies at fixed viewing distance and eccentricity but different speeds.

Results

We measured cut-off frequencies for a triangular speed modulation function with a modulation depth of 80% at a *fixed* 10 deg eccentricity and fixed viewing distance (240 cm), but at different speeds v_0. Because sweep-time t_0 was constant (1.25 s), the sweep-length d_0 was directly proportional to the dot speed.

— Figure 4 about here —

Cut-off modulation frequencies are shown in Fig. 4 as a function of the average speed v_0 (at fixed eccentricity!). As expected, we find a dependence of cut-off modulation frequency on speed.

Discussion

We fitted the dependence of the cut-off frequency on speed to a power function (see Eq. 4). The exponent α that fits the data of Fig. 3 best is estimated to be $\alpha = 0.25$ for HS and $\alpha = 0.30$ for AT. The *absolute* values of the cut-off frequencies at a fixed speed in Fig. 4 can be compared with Fig. 3. The cut-off frequencies of this experiment (measured with 80% modulation amplitudes) are roughly 0.8 times the cut-off frequencies in Fig. 3 (measured for 100% modulation amplitudes). This can be explained by the fact that the filtered modulation signal is proportional to the modulation amplitude dv_x times an attenuation function (see Model Section). This filtered signal has to exceed a certain internal threshold for detection to take place. Thus a higher modulation amplitude yields higher modulation frequencies.

Note, that the cut-off frequencies at the lowest average speeds in Fig. 3 were measured at much smaller eccentricities than the 10 deg eccentricity for this experiment. Thus, it seems that eccentricity of presentation is of small relevance to speed modulation detection, even for speeds that barely exceed the motion *detection* threshold at the eccentricity of presentation.

Finally, it is of interest to note that human modulation detection sensitivity at the lowest speed tested (0.5 deg/s at 10 deg eccentricity) is excellent when expressed in terms of the spatial excursions of the modulated motion path from the average motion path at average dot speed v_0. These spatial excursions did not exceed 0.9 arcmin, which is approximately the hyperacuity threshold that was found at this eccentricity for static stimuli with an explicit nearby spatial reference available (Westheimer, 1982)!

Direction Modulation Detection

In this section, we study velocity vector modulations *orthogonal* to the average velocity vector (direction modulations) resulting in curved trajectories. Instead of measuring speed modulation thresholds, we measure direction modulation thresholds. The precise generation of the direction modulation functions is specified in the method section, as is the definition of direction modulation thresholds. Otherwise, the procedure and organization of these experiments are quite similar those of the speed modulation detection experiment described above.

Direction Modulation Detection Dependence on Modulation Frequency

Results

In Fig. 5, we present direction modulation thresholds as a function of modulation frequency for two different direction modulation functions.

— *Figure 5 about here* —

Circles represent sinusoidal direction modulations $\Omega(t)$, squares block shaped direction modulations $\Pi(t)$. Since direction modulation thresholds were very similar across the main observers, we presented the averaged thresholds for observers AT and HS. The parameter settings for different velocities and functions are listed in Table 4.

— *Table 4 about here* —

Direction modulation detection thresholds for an average speed $v_0 = 1$ deg/s and sinusoidal modulation functions (small filled circles in Fig. 5) are approximately invariant ($d\theta = 10.2-12.3$ deg) for the range of frequencies tested (0.25-1 Hz).

At high frequencies, thresholds rise strongly as a function of frequency ω for both sinusoidal and block shaped direction modulation functions. The average ratio of threshold amplitudes for sinusoidal and block shaped modulation functions at high frequencies $\omega \geq 2$ Hz is 1.26.

Discussion: Threshold Invariance at Low Modulation Frequencies

The threshold invariance at low frequencies indicates that human direction modulation detection is based on the *amplitude* of the direction modulation function and *not* on its temporal *derivative* (directional acceleration); direction modulation amplitude is independent of modulation frequency, whereas the derivative is linear with modulation frequency. Therefore, a detection based on directional acceleration would yield increasing thresholds for decreasing low modulation frequencies, which is *not* observed.

On the other hand, direction modulation detection might be based on the magnitude of vertical dot position (the temporal *integral* of the direction modulation function). However,

a constant vertical position threshold (relative to the mean position) would yield decreasing threshold direction modulation amplitudes for decreasing modulation frequencies in Fig. 5, which is also *not* observed.

At 1 Hz modulation frequency, the 6.6 deg threshold direction amplitude at $v_0 = 2$ deg/s is smaller than the 10.2 deg threshold found at $v_0 = 1$ deg/s. Although here we compare only two data points we speculate that the asymptotic level is speed dependent. Increasing direction modulation thresholds at low speeds are not surprising given the relatively low sensitivity of the human motion system at low velocities (DeBruyn and Orban, 1988).

In conclusion, the observed asymptotic behavior of direction modulation thresholds as a function of modulation frequency support the hypothesis that detection is based on the *amplitude* of the direction modulation function. The observed absence of mechanisms tuned to visual acceleration seems consistent with a study on motion after effects (MAE) by Schwartz and Kaufman (1987), who reported that "there is no MAE specific to adaptation for changing directions as distinct from simple motion".

Discussion: Wave Forms

Because the fundamental frequency of sinusoidal functions has a smaller amplitude (a factor of $\pi/4$) than that of block functions (at a given modulation amplitude), thresholds are expected to be a factor of $4/\pi = 1.27$ higher for sinusoidal than for block functions at relatively high modulation frequencies. The average ratio (1.26) of threshold amplitudes for sinusoidal and block shaped modulation functions found for the frequency range ≥ 2 Hz, is in good agreement with this ratio, suggesting a relatively low cut-off frequency. The strong increase of direction modulation thresholds for both sinusoidal as for block functions, suggests a temporal frequency limit for the direction detection system of approximately 2 Hz. We found a similar temporal limit for speed modulation detection.

Discussion: Direction Modulations v. Speed Modulations

To compare the (absolute) direction modulation detection thresholds $d\theta$ with the Weber fractions for speed modulations, we use the following Weber fraction W_d for direction modulations: $W_d = dv_y/v_0 = \tan d\theta$. Now, both Weber fractions $W_d = dv_y/v_0$ and $W_v = dv_x/v_0$ are elegantly expressed in terms of velocity vector modulations and can be compared.

For example, one can compute that the Weber fraction W_d for block shaped direction modulation detection at 1 Hz and speed $v_0 = 1.7$ deg/s is approximately 9%. This is about a factor 2 *lower* than the Weber fraction $W_v = 17\%$ for block shaped *speed* modulation detection. Furthermore, if one takes into account the effects of probability summation (or variance detection of the velocity vector modulation), very similar ratios (a factor of 2) for speed and direction modulation thresholds can be shown to hold for other modulation functions used (see general discussion section).

The fact that the human visual motion system is more sensitive to direction than to speed is a well-known phenomenon in motion *discrimination* experiments, where Weber fractions for speed discrimination are typically *twice* the Weber fractions for direction discrimination

(Nakayama, 1985b; DeBruyn and Orban, 1988). Thus, —although ***absolute*** Weber fractions for ***modulation detection*** are a factor 3-4 higher than Weber fractions for ***motion discrimination*** experiments— the ***ratio*** of Weber fractions for speed and direction modulations is very similar (about a factor of 2) in both types of experiments.

The invariant ratio of sensitivity to speed and direction reveals a fundamental characteristic of the human visual motion system. Let's consider a motion system consisting of an ensemble of Reichardt correlators that have spatially rotational-invariant prefilters (*i.e.*, spatial input filters that are *not* orientation selective) (see Glünder, 1990). For such a motion system, the higher sensitivity to direction than for speed is likely to be related to the fact that in the motion direction *both* spatial and temporal prefilters contribute to a broadening of the ensemble response correlation peak, whereas in the orthogonal direction the width of the correlation peak is determined solely by the spatial prefiltering.

Discussion: Low-pass Temporal Filtering at High Frequencies

We proposed a low-pass temporal filter to explain the cut-off frequency for speed modulation detection. The temporal limit found for speed modulations (2 Hz) was similar to the cut-off frequency of approximately 2 Hz observed here for direction modulation detection. In the Model Section, we model this temporal behavior of direction modulation detection analogous to the way we modeled that of speed modulation detection: a temporal filtering of the direction modulation signal which corresponds to a temporal filtering of the speed signal $v_y(t)$ orthogonal to the mean velocity vector (see Eq. 3).

Spatial Cues at Very High Modulation Frequencies

Motivation

For modulation frequencies that far exceed the temporal limit (2 Hz) for detection of modulation by the motion system, thresholds are determined by other cues outside the motion system, such as spatial cues. Although these thresholds do not reveal the characteristics of the motion system, they are of interest.

Results

We measured direction modulation thresholds at 8 Hz and 100 Hz modulation frequency for a 1.7 deg/s speed frequency. Thresholds were found to level off for this frequency range, suggesting that observers made use of weaker *spatial* cues to detect modulations, which were independent of modulation frequency.

The threshold for the remaining *spatial excursion* cue at high frequencies is a measure for the dynamic spatial acuity *orthogonal* to the motion direction of a moving dot. We found it to be about 8 arcmin for our observers, which is a few times the 2-3 arcmin acuity limit for static stimuli (Wertheim, 1894) at the eccentricity used in this experiment (3/4 deg).

Pulse shaped Direction Modulation Functions

Motivation

Here we make a short study of pulse shaped *direction* modulation functions $\delta(t)$, yielding square wave vertical *position* modulation functions $\Pi(t)$ (see Fig. 1). Detection of pulse shaped direction modulation functions is interesting because the corresponding vertical velocities are too high to be sensed by the motion system. Hence, modulation detection must be based on cues outside the motion system. Therefore, the detection of pulse shaped modulation functions is likely to make use of a spatial cue: the spatial excursion from the mean vertical position.

Filtered pulse shaped direction modulation functions have constant amplitudes for low modulation frequencies (when the response functions in the time domain for consecutive pulses are well-separated), yielding constant spatial excursions from the mean vertical position. Therefore, if the spatial excursion is the cue for modulation detection, thresholds are expected to be independent of modulation frequency for low frequencies. We tested this prediction by measuring modulation detection thresholds (expressed as spatial excursions) as a function of modulation frequency for pulse shaped modulation functions $\delta(t)$.

Results

The parameter settings for this pulse detection experiment are listed in Table 4.

— *Figure 6 about here* —

Figure 6 shows that the spatial excursions yielding threshold performance are approximately invariant (2.3-2.8 arcmin) for the range of modulation frequencies tested (1-4 Hz). Similarity in performance for subjects AT and HS allowed averaging over these observers.

Discussion

The finding that the spatial thresholds at the 1 Hz modulation frequency are slightly higher than those for 2 Hz may be explained by probability summation or by the nature of variance detection. At the fixed stimulus presentation time used there are twice as many 'events' (pulses) at a 2 Hz than at a 1 Hz modulation frequency. Consequently, probability summation or variance computation can take place across more events yielding lower thresholds. Obviously, there is a limit to summation such that no improvement occurs at even higher modulation frequencies.

Note that the spatial modulation thresholds (spatial excursions) (2-2.5 arcmin) are considerably lower than the dynamic *acuity* (8 arcmin) determined earlier in this paper at the same speed (1.7 deg/s). However, they are still higher than the hyperacuity thresholds (0.2-0.5 arcmin) that have been measured for stationary spatial configurations at similar eccentricities (Westheimer, 1982). A more elaborate experiment that we performed on pulse shaped *speed* modulation functions will be presented elsewhere.

The fundamental frequency components of block shaped and triangular shaped *position* modulation functions (at a given modulation amplitude) have relative amplitudes $\pi/2$. With a low cut-off frequency of 1 Hz, these fundamental frequencies are expected to dominate the detection thresholds at high modulation frequencies. Therefore, we expect relative thresholds of $2/\pi \approx 0.64$. This can be verified by comparing thresholds for block shaped position functions (see pulse shaped *direction* modulations in Fig. 6) with the triangular shaped position functions (see block shaped direction modulations in Fig. 5) at, for example, 4 Hz modulation frequency. This average ratio across observers is 0.67 supporting our claim that we deal with positional cues and a low cut-off frequency such that the fundamental frequencies dominate detection of high frequency modulation functions.

Model

We present a model for the detection of velocity vector modulations in the human visual system. The model consists of two stages. The first stage is a nth order low-pass temporal filter that operates on the velocity vector modulation function. This filter is characterized by its order n and a characteristic time constant τ. The second stage is a decision stage based on the filtered modulation function.

In this section, we will show that our data provide strong experimental evidence that the decision stage is a *variance detection* stage. The single parameter that specifies the variance detection stage is a variance threshold and can estimated from the data. Furthermore, having knowledge about the decision stage, we can estimate the parameters that characterize the first (low-pass temporal filter) stage.

The Decision Stage: Variance Detection

The amplitude detection thresholds presented in this paper have been interpolated based on detection probabilities as measured using a method of constant stimuli. Direction thresholds are the modulation amplitudes at threshold performance (80% correct answers) and thus form one parameter to characterize the full psychometric functions available. However, the *shape* of the psychometric functions reveal the parameter used in the decision stage (e.g., velocity, squared velocity, etc). Therefore, it is of interest to examine the shape of the psychometric functions.

Due to the noise, associated with the stochastic (binomial) nature of the observer decision process, we need a large number of elementary decisions for each modulation amplitude in order to discriminate small shape differences of different psychometric functions. Therefore, we averaged psychometric functions, for all parameter settings used in the modulation detection experiments described above, as a function of *normalized* velocity vector modulation

amplitudes ξ. With a normalized vector modulation amplitude ξ, we mean a modulation amplitude for a given parameter setting, divided by the modulation amplitude yielding threshold performance for that particular setting: $\xi = (dv/v_0)/W_v$. Assuming that velocity vector modulation detection is ruled by a single detection process, psychometric functions for all parameter settings will be identical if plotted as a function of normalized amplitudes!

The resulting psychometric curve that describes all experiments reported in this paper is shown in Fig. 7.

— Figure 7 about here —

The horizontal axis in Fig. 7 represents normalized modulation amplitudes ξ. The ordinate represents the percentages correct for a small range of normalized modulation amplitudes clustered around a range of plotted normalized amplitudes of the data points. Half the length of the shown error bar for a normalized amplitude i corresponds to the square-root-variance (σ_i) of the binomial probability distribution for that point: $\sigma_i = \sqrt{p_i(1-p_i)}/\sqrt{n_i}$, with p_i the ordinate for modulation i, and n_i the number of elementary observer decisions. Each individual data point is based on about 700 elementary (yes/no) observer decisions ($n_i = 700$).

We find an excellent fit of this psychometric function using a *standard* error function or (scaled) standard normal distribution $Erf(z)$:

$$Erf(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\gamma z} e^{-(\frac{i^2}{2})} di, \tag{5}$$

that has the *square* normalized modulation amplitude ($z = \xi^2$) as its argument (χ^2=5.2 with 7 degrees of freedom). The scale factor γ is *constant*: $\gamma = 0.84$, and causes $Erf(z)$ to be 80% for $z = 1$ (a constraint, set by our threshold definition). Note that the function $Erf(\xi^2)$ used for the fit has no free parameters! The evidence for this particular shape of psychometric function is very strong since a fit with functions $Erf(\xi)$ or $Erf(\xi^3)$ yields unacceptable Chi-square values of $\chi^2 = 152$ and $\chi^2 = 55$ respectively. Thus, assuming that a standard error function is a valid description of the psychometric function associated with the final observer decision process (Green and Swets, 1966), the empirical function shown in Fig. 7 strongly suggests that modulation detection is based on the *square* modulation amplitude, *i.e.*, the *variance* of the (temporally filtered) velocity vector modulation.

A variance detection process is certainly not exclusive for velocity vector modulation detection and has, for example, also served to explain human sensitivity to temporal fluctuations in the *luminance* domain (Rashbass, 1970; Koenderink and van Doorn, 1978). The choice of variance detection for these visual tasks is not surprising since variance detection has been shown to be optimal from a statistical point of view in a number of instances in which the visual system is *uncertain* about some aspects of the stimulus to be presented (Green and Swets, 1966), *e.g.*, the phase of periodical modulation functions. This uncertainty may force

the observer to use the autocorrelation (variance) of the velocity modulation signal, instead of the more efficient cross correlation of the modulation signal received with the signal expected (Green and Swets, 1966; Burgess and Ghandeharian, 1984).

We argued that the uncertainty of the observer about the phase of the modulation function accounts for variance detection and thus for our finding that speed *modulation detection* thresholds (17%) are much higher than speed *discrimination* thresholds (6%). However, an alternative explanation for variance detection is in terms of the difficulty of the visual system in segmenting modulated motion paths in temporal segments of different velocities in the absence of cues other than in the motion dimensions (Snowden and Braddick, 1991). To decide on this issue, experiments are useful in which the observer is provided with explicit cues that define the phase of the motion modulation signal and/or allow for a segmentation of the stimulus.

Interestingly, such experiments have been performed in another domain. In stereo vision (disparity processing), for example, the absence of an explicit segmentation cue (*i.e.*, defined outside the stereo domain) can lead to dramatic increases in the disparity discrimination threshold (McKee, 1983; Fahle and Westheimer, 1988). In the domain of two-dimensional luminance pattern perception, on the other hand, an explicit segmentation cue has been observed to force the visual system in a processing mode yielding higher discrimination thresholds than obtained without segmentation cues (Watt, 1985)!

Rashbass (1976) has shown that a *variance detection* process based on the *square* modulation amplitude yields identical detection performance to an alternative detection model consisting of *peak detection* based on the *linear* modulation amplitude, with a detection probability summation that is governed by a psychometric function similar in form to the empirical curve shown in Fig. 7.

We have shown that modulation detection is accurately described by a variance detection, based on a low-pass transformation of the velocity vector function, that is, velocity vector modulations are detected whenever the variance (after filtering) exceeds a certain internal threshold. The internal thresholds are most easily derived from threshold modulation amplitudes for block shaped modulation functions at a low frequency, when the modulation function is nearly unaffected by the low-pass filter. This is because the variance of block shaped functions is equal to the square modulation amplitude. Therefore, the internal threshold for variance detection is equal to the square of the threshold modulation amplitude measured for these functions. For example, the minimum standard deviation (square root variance) yielding modulation detection is $W_v = 17\%$ for speed modulations (see Fig. 2) and $W_d = 9\%$ for direction modulations (see Fig. 5).

The Filter Stage: A Second Order Low-pass Filter

The data presented in this paper suggested that the detection of velocity vector modulation functions is based on a temporal low-pass filtered version of the true (physical) stimulus velocity vector function. The effects of such a low-pass temporal filter on detection performance depends on the type of decision stage that follows temporal filtering. Now we have specific knowledge

about this decision stage (see above), we can estimate the temporal characteristics of the first (filter) stage.

We model the temporal low-pass filter as a standard temporal n-th order low-pass filter and estimate its order and its characteristic time constant from the dependencies of modulation thresholds on modulation frequency

Method

A standard temporal n-th order low-pass filter has a pulse response function:

$$g(t) = \frac{1}{\tau(n-1)!}\,(\frac{t}{\tau})^{n-1}\,e^{-t/\tau} \quad (n > 0), \tag{6}$$

and a transfer function $\tilde{g}(\omega)$ of modulation frequency ω:

$$\tilde{g}(\omega) = [1 + (2\pi\omega\tau)^2)]^{-n/2}. \tag{7}$$

This filter reduces the amplitude of the modulation functions. For example, a sinusoidal modulation function with frequency ω and amplitude A passing the filter $g(t)$ will have a reduced amplitude $A\tilde{g}(\omega)$. The detection of such a modulation signal takes place in a variance detection stage as described above. At threshold, the variance of this filtered sinusoidal signal is equal to a threshold variance σ_0^2: $A^2.\tilde{g}^2(\omega)/2 = \sigma_0^2$ (recall that the variance of a sinusoidal signal equals half its square amplitude).

This threshold σ_0^2 can be estimated from an empirical modulation detection threshold $W(\omega_0)$ at a low modulation frequency $\omega_0 \ll 1/(2\pi\tau)$ for which $\tilde{g}(\omega_0) \approx 1$. For example, for block shaped modulation functions the variance of the function at threshold amplitude $W(\omega_0)$ is exactly $W^2(\omega_0)$. Thus, $\sigma_0^2 = W^2(\omega_0)$ for block shaped modulation functions.

To estimate the time constant τ, we consider the threshold amplitudes at modulation frequencies ω_1 and ω_2 for which the the higher order spectral components of the modulation function can be ignored. In those cases, the modulation functions are approximated by their fundamental (sinusoidal) components. For example, for block shaped functions with (threshold) amplitude $W(\omega_i)$, the amplitude of the fundamental sinusoidal component is $\frac{4}{\pi}W(\omega_i)$. At threshold, the variance of the filtered fundamental equals the detection threshold σ_0^2:

$$\frac{1}{2}[\frac{4}{\pi}W_v(\omega_i)]^2[1 + (2\pi\omega_i\tau)^2)]^{-n} = \sigma_0^2. \tag{8}$$

For frequencies $\omega_i \gg (2\pi\tau)^{-1}$, we could use the asymptotic behavior $\tilde{g}(\omega_i) \approx (2\pi\omega_i\tau)^{-n}$ to estimate τ and n analytically from two data points at high frequencies. However, we can not use this approximation a priori and use a numerical approach. First, we rewrite Eq. 8 to solve τ^2 as a function of n and ω_i:

$$\tau^2(\omega_i, n) = \frac{1}{(2\pi\omega_i)^2}[(\frac{\sqrt{8}W_v(\omega_i)}{\pi\sigma_0})^{\frac{2}{n}} - 1]. \quad (9)$$

The time constant τ is constant. Thus, $\tau(\omega_i, n)$ is expected to be the same for any two different (sufficiently high) modulation frequencies ω_1 and ω_2. Hence, the order n of the filter is the solution of the equation:

$$\tau(\omega_1, n) = \tau(\omega_2, n). \quad (10)$$

We chose the lowest n for which $|\tau(\omega_1, n) - \tau(\omega_2, n)|/[\tau(\omega_1, n) + \tau(\omega_2, n)]$ is smaller than 20%. The time constant τ is taken to be the average of the two values $\tau(\omega_1, n)$ and $\tau(\omega_2, n)$ at this n.

Estimation of Filter Parameters for Speed Modulation Thresholds

Consider the speed modulation thresholds for block shaped functions in Fig. 2. We use the threshold at low modulation frequency $\omega_0 = 1$ Hz for the estimation of σ_0: $\sigma_0 = W_v(1) = 17\%$. Furthermore, we use the two thresholds at high modulation frequencies $\omega_1 = 4$ Hz and $\omega_2 = 8$ Hz with thresholds $W_v(\omega_1) = 30\%$ and $W_v(\omega_2) = 81\%$.

Using the method described above, we find that a value $n = 2$ and a time constant $\tau = 33$ ms adequately model the dependence of speed modulation detection thresholds $W_v(\omega)$ as a function of modulation frequency ω. In fact, we found $\tau(\omega_1, 2) = 30.6$ ms and $\tau(\omega_2, 2) = 36.1$ ms.

With $\tau = 33$ ms, this second order low-pass filter corresponds to a value of approximately 90 ms for the full width at half maximum (FWHM) of the pulse response of the speed integration filter yielding an integration (smoothing) of the physical speed signal in the human visual system within roughly a 100-140 ms temporal window.

Estimation of Filter Parameters for Direction Modulation Thresholds

Similar to the previous section, we estimated the order n and time constant τ from the direction modulation thresholds for block shaped functions as presented in Fig. 5. We used the threshold $W_d(\omega_0) = 8.7\%$ for $\omega_0 = 1$ Hz, and $W_d(\omega_1) = 11.4\%$ ($\omega_1 = 2$ Hz) and $W_d(\omega_2) = 24.9\%$ ($\omega_2 = 4$ Hz).

Using the method described above, we find that a value $n = 2$ and a time constant $\tau = 42$ ms adequately model the dependence of direction modulation detection thresholds $W_d(\omega)$ as a function of modulation frequency ω. In fact, we found $\tau(\omega_1, 2) = 34$ ms and $\tau(\omega_2, 2) = 49$ ms.

This time constant for the low-pass filtering of the direction modulation signal is only slightly higher than the time constant (τ=33 ms) estimated for the low-pass filtering of the speed modulation signal. However, as we showed in Fig. 3, this small discrepancy may be a consequence of the different speed ranges used for the determination of the temporal filter characteristics for direction and speed modulations.

The overall similarity of the characteristics of human detection of speed and direction modulations and the (near) equality of the integration time-constants derived strongly suggests a detection system that monitors the full (temporally filtered) velocity *vector*.

General Discussion

Evidence for *indirect* optic acceleration detection

We presented a study of human sensitivity to optic acceleration and have been unable to find any evidence for a visual mechanism that *directly* detects optic acceleration, *i.e.*, the temporal derivative of the velocity vector modulations. Instead we find strong evidence that modulation detection is based on the *amplitude* or modulation depth of a temporally filtered velocity vector modulation signal. The temporal characteristics of the temporal filter are adequately described by a second order low-pass filter with a time constant $\tau \approx 40$ ms. Effectively, this filter corresponds to a temporal integration of the velocity signal of at least 100 ms. This is consistent with the upper temporal limit of about 100 ms for the integration of velocity information (improving signal-to-noise ratios) in motion discrimination experiments (DeBruyn and Orban, 1988; Snowden and Braddick, 1991). Thus, the lower and upper limits for temporal integration in the human visual motion system are equal, suggesting a single hard-wired temporal filter in the motion processing system. This view is further supported by the close quantitative correspondence between the increase of cut-off frequency with speed (as reported here for motion modulation detection) and the decrease of temporal integration time with speed found in motion discrimination studies (van Doorn and Koenderink, 1982, 1985).

This leads to the intriguing question: which stage in the stream of visual motion processing accounts for the characteristic temporal filtering found in our experiments?

Temporal Filtering: Mechanistic Considerations

A functional description of the phenomenology of our experiments consists of a temporal integration of an unsmoothed internal representation of the true velocity signal (Fig. 8a). At this point we will try to link this functional description to an actual implementation in the visual system in terms of well-known motion detection mechanisms.

— *Figure 8 about here* —

An abstract description in terms of a smoothed motion signal does not necessarily mean that the visual system actually extracts an exact (unfiltered) velocity signal to subsequentially low-pass filter it in time. In fact, the following rhetorical questions make such an implementation

unlikely: (1) How does the visual system arrive at the representation of the true (unfiltered) velocity in the first place? (2) If such a representation exists, should this signal be low-pass filtered given the great advantages of having access to a velocity signal with high temporal resolution (Arnspang, 1988; Glünder, 1990)? Because of the above puzzles, we believe that the temporal integration is inherent to the mechanism that arrives at a velocity representation and that it takes place effectively before the final estimate of the velocity vector, akin to the scheme of Fig. 8b.

We illustrate some of the possible stages of temporal filtering by adopting a specific but plausible basic motion detector: the Reichardt-correlator (van Doorn and Koenderink, 1985), see Fig. 9.

— *Figure 9 about here* —

A plausible implementation of such a correlator typically contains three temporal filtering stages:

1. A temporal prefilter $f(t)$ for each input line.
2. A temporal delay filter (with time constant τ) in one of the input lines.
3. A temporal low-pass filter $\int_T$ of the correlator output.

We will discuss each of these filters as candidates to account for the temporal low-pass filtering of the velocity vector modulation functions found in our experiments.

Temporal Low-pass Filtering of Correlator Output

Intuitively, it is tempting to associate the psychophysically observed integration of velocity with the temporal integration $\int_T$ of the correlator output. Perhaps surprisingly, however, temporal filter $\int_T$ is *not* equivalent with a temporal integration of the modulation of speed or direction. To show this, we will consider an ensemble of motion detectors (Reichardt correlators), ideally tuned to a continuum of velocities. To substantiate our point we will focus on speed modulation functions. Assuming that detectors tuned to identical velocities are pooled (Glünder, 1990), this ensemble can be parameterized by tuning velocity v_t only. At time t, the moving target has velocity $v(t)$ and will thus activate only detectors with a tuning velocity $v_t = v(t)$. Therefore, the type of activated detectors (parameterized by v_t) within the ensemble will vary in time, yielding time dependent *ensemble activation profiles.* For example, ensemble activation profiles for a triangular speed modulation function are given in the left upper corner of Fig. 10.

— *Figure 10 about here* —

In this figure the type of detector (parameterized by its tuning speed v_t, along the vertical axis) that is activated by the moving dot is given as a function of time t (along the horizontal axis). When the detectors are very sharply tuned, only one type of detector is active at a time, dependent on the speed of the moving dot. At a particular moment in time, we can walk along the vertical axis and find which detectors are active as a function of their tuning speed (the ensemble activation profile). Because the dot moves with a single speed at a each moment in time, the ensemble profile is a single pulse that shifts along the vertical axis in time. With sharply tuned detectors, the ensemble activation profile is a perfect copy of the physical speed modulation signal $v(t)$, and is thus triangular in time (see upper left corner of Fig. 10).

Now let's consider the temporal filter $\int_T$ that integrates the output of the standard motion detectors in time yielding an integration of the ensemble activation profile along the horizontal time axis. The resulting horizontally blurred activation profile is shown in the upper right corner of Fig. 10. To make our argument as strong as possible, we assumed that the temporal integration takes place within a temporal window that exceeds the period of the modulation function a few times such that it flattens the profile.

Figure 10 also shows the ensemble activation profiles for a uniform (unmodulated) speed function *before* (bottom left) and *after* temporal integration (bottom right). Obviously a constant profile in time is invariant under temporal integration.

As a result of temporal integration $\int_T$ (blurring along the horizontal time axis), both the ensemble activation profiles for modulated and unmodulated velocity functions are *constant* in time. However, the shape of the profiles for the modulated speed function (upper right) and that for the unmodulated speed function (bottom right) differ strongly even for inifinite blurring.

A true integration of the speed modulation signal, however, would blur the ensemble profile along the vertical (speed) axis yielding blurred profiles that become indiscriminable for infinite blurring.

The above reasoning shows that temporally filtering the speed signal is not equivalent to temporally filtering the output of motion detectors (correlators). That is, blurring is the speed dimension is generally not equivalent to blurring in the time dimension. We suggest that the psychophysically observed integration of the speed signal must be inherent to a processing stage which comes *before* the correlation stage.

Temporal Pre-filter

The shown temporal low-pass characteristic for modulation detection might be inherent to the temporal filter $f(t)$ at the input of a standard motion detector (see Fig. 9). The reasoning would be in terms of 'window of visibility' arguments as used to explain the perceived equality of apparent motion with real motion at adequate sampling frequencies (Watson *et al.*, 1986; Burr *et al.*, 1986). However, in order to account for the low-pass filter characteristics for modulation detection, we have to assume upper temporal cut-off frequencies of 4-8 Hz. At these unrealistic temporal cut-off frequencies, the motion system has not even reached optimal sensitivity (Burr and Ross, 1982)! Consequently, the temporal pre-filter explanation

is implausible.

Temporal Delay Filter

We suggest that the temporal delay filter (see Fig. 9), is the most plausible candidate to account for low-pass transformations of modulation functions. The standard motion detector as sketched in Fig. 9 is optimally activated if an object traverses the spatial interval between the front end receptors of its two input lines in the finite delay time τ. The speed of the object may vary during its trajectory, as long as the above constraint is satisfied. Intuitively, this results in a temporal averaging (or temporal integration) of the speed function.

Glünder (1989) has recently presented an interesting mathematical analysis on this issue. His study focused on the question of how velocity estimates through an ensemble of standard motion detectors depend on the spatial object function and on the impulse response function of the delay filter of the detector. For an ensemble of bilocal correlators tuned to a continuum of velocity vectors, he showed that the estimated velocity function is the result of the convolution of the true (physical) velocity vector function with a time-invariant kernel which only depends on the integral function of the impulse response function. Hence, the estimated velocity vector function is independent of the spatial object function.

Glünder's proof strongly supports our view that the phenomenological description of our results, in terms of low-pass temporal filtering of the velocity vector function, corresponds with a plausible implementation in terms inherently non-ideal (realizable) band-pass delay filters in correlator detectors. Following this hypothesis, the cut-off frequencies for modulation detection depend inversely on delay value τ: for longer delay values, the width of convolution kernel increases and yields stronger temporal blurring. Our finding that cut-off frequencies ω_c are slightly dependent on speed v_0 ($\omega_c(v_0) \propto v_0^{0.35}$) thus lead to the conclusion that delay values τ are speed dependent. Cut-off frequencies ω_c are expected to be inversely dependent on correlator delay τ. This conclusion corresponds closely to the empirical power function reported by van Doorn and Koenderink (1982): $\tau \propto v_0^{-0.40}$.

Speed Dependence of Cut-off Frequencies

The dependence of cut-off frequency on speed we observe in Fig. 3 is consistent with what has been reported in both psychophysical and electrophysiological literature: higher velocities correspond to somewhat faster detectors.

However, this finding is at odds with the speed dependence of temporal velocity resolution obtained by Snowden and Braddick (1991). They found that the cut-off frequency for speed modulation detection *decreases* with increasing velocity in their experimental set-up. This issue remains to be resolved by further experimentation. The most noticeable difference between our experiments and those of Snowden and Braddick concerns the spatial nature of the stimuli used. Snowden and Braddick used a spatially extended random dot pattern centered at the fovea, whereas we used a localized target moving at a trajectory with an approximately constant eccentricity in our case.

Relation to Sampled (Apparent) Motion Experiments

Apparent motion differs from continuous (smooth) motion because it is characterized by a speed function that is modulated in time. Apparent motion is thus a special case of the modulation functions examined in this paper. We will discuss two studies (Watson *et al.*, 1986 and Burr *et al.*, 1986) that reported on the minimum temporal sampling frequency yielding perceptual equivalence of apparent and real motion and compare them with our study on the upper cut-off frequencies for velocity modulation detection.

In Watson *et al.*'s stroboscopic paradigm the time-dependent speed is an ill-defined signal, but periodic with a frequency equal to the strobe frequency. They find minimal sampling frequencies (yielding perceptual equivalence of stroboscopic and real motion) that are much higher ($>$ 30 Hz) than the cut-off frequencies for velocity modulation detection obtained in this study. However, this is probably due to the strong luminance cue at low velocities.

In Burr *et al.*'s 'sample and hold' paradigm the moving dot is visible all the time and displaced stepwise in time. The time-dependent speed function $\Upsilon(t)$ is now well-defined (see Fig. 1) and can be compared with the block modulation functions used in our experiments with 100% modulation amplitude. Fig. 1 shows that 'sample and hold' speed modulation function $\Upsilon(t)$ differs *only* in duty cycle from the block modulation function $\Pi(t)$ at 100% modulation amplitude as used in our experiments. Furthermore, functions $\Upsilon(t)$ differ only in peak width from the triangular functions $\Lambda(t)$ at 100% amplitude. However, the cut-off frequencies found in this paper (approximately 4 Hz for triangular modulation functions at $v_0 = 2$ deg/s, see Fig. 3) differ by at least a factor of 4 from the minimum sampling frequencies (15-40 Hz, dependent on the drift rate) found by Burr *et al.* for this particular speed. We offer a number of explanations for this apparent discrepancy:

1. For a given average speed v_0, the amplitude of the fundamental frequency in the 'sample and hold' motion modulation function $\Upsilon(t)$ is much larger than for the triangular speed modulation function $\Lambda(t)$ we used to obtain Fig. 3 (the ratio equals $\pi^2/4 \approx 2.5$). This allows for higher cut-off frequencies in Burr's paradigm.

2. Burr *et al.* do not mention visual fixation. The strong dependence of detection performance on pursuit eye movements was discussed in the Method section. Our observers reported to have *no difficulties* in detecting modulations when tracking the moving dot for conditions where detection *failed* under visual fixation.

3. We mentioned before that for modulation frequencies that far exceed the temporal limit (2 Hz) for the detection of modulations by the motion system, thresholds are determined by cues *outside* the motion system, such as spatial cues. Thus it may be that experiments on the equivalence of apparent and real motion do not exclusively reveal the structure of the visual motion system.

Relation to Experiments with Controlled Eye Movements

Results have been reported on frequency limits for velocity modulation detection when a *moving* reference is provided (Funakawa, 1989) (contrary to our *stationary* fixation dot). Providing a moving reference leads to cut-off frequencies ($\approx$ 25 Hz) that are considerably higher than those obtained in our study. Interesting as these results are, we believe them to be indicative of the temporal resolution of visual subsystems concerned with the *spatial* analysis of moving patterns, and not with the determination of velocity as such. We believe both types of experiments (and visual subsystems) should be clearly distinguished (although, of course, they may be intimately intertwined).

Conclusion

In conclusion, human detection of velocity vector modulations is *not* based on optic acceleration (the temporal derivative of the velocity modulation function $v(t)$). The data presented in this paper strongly support the view that modulation detection consists of a *variance detection process*, based on the *magnitude* of a low-pass filter transformation of the true modulation function $v(t)$. Effectively, the motion system integrates the velocity vector modulation signal for about 100 ms over time.

These results put severe constraints on viable theories aiming to explain human capacities in the extraction of 3D environmental information from motion parallax cues (Nakayama, 1985a).

Acknowledgement

This research was performed at the Institute for Perception TNO, Soesterberg, The Netherlands. The research of Peter Werkhoven was supported by the USAF Life Science Directorate, Visual Information Processing Grant 88-0140. Herman Snippe was supported by the InSight project of the ESPRIT Basic Research Actions of the European Community. Alex Toet was supported by NATO grant CRG 890970.

References

[1] Arnspang J. (1988) Optic Acceleration, *Proceedings ICCV*, 364–373, 1988.

[2] Burgess A and Ghandeharian H (1984) Visual signal detection. I. Ability to use phase information, *J. of the Optical Society America A* **1**, No. **8**, 900–905.

[3] Burr D.C. and Ross J. (1982) Contrast sensitivity at high velocities, *Vision Research* **22**, 479–484.

[4] Burr D.C., Ross J. and Morrone M.C. (1986) Smooth and sampled motion. *Vision Research* **26**, 643–652.

[5] DeBruyn B. and Orban G.A. (1988) Human velocity and direction discrimination measured with random dot patterns, *Vision Research* **28**, 1323–1335.

[6] Doorn A.J. van and Koenderink J.J. (1982) Temporal properties of the visual detectability of moving spatial white noise, *Experimental Brain Research* **45**, 179–188.

[7] Fahle M. and Westheimer G. (1988) Local and global factors in disparity detection of rows of points, *Vision Research* **28**, 171–178.

[8] Funakawa M. (1989) Spatio-temporally averaged positions of moving objects, *Spatial Vision* **4**, No. **?**, 275–285.

[9] Glünder H. (1990) Correlative velocity estimation: visual motion analysis, independent of object form, in arrays of velocity tuned bilocal detectors, *J. of the Optical Society America A* **7**, No. **2**, 255–263.

[10] Green D.M. and Swets J.A. (1966) Signal detection theory and psychophysics, New York: Wiley.

[11] Hick W.E. (1950) The threshold for sudden changes in the velocity of a seen object, *Quarterly Journal of Experimental Psychology* **2**, 33-41.

[12] Julesz B. (1971) Foundations of Cyclopean Perception, Chicago: University of Chicago Press, 1971.

[13] Koenderink J.J. (1986) Optic flow, *Vision Research* **26**, 161–180.

[14] Koenderink J.J. and Doorn A.J. van (1978) Detectability of power fluctuations of temporal visual noise, *Vision Research* **18**, 191–195.

[15] Koenderink J.J., Doorn A.J. van and W.A. van de 1985 (Grind) 1985, Spatial and temporal parameters of motion detection in the peripheral visual field *J. of the Optical Society America A* **2**, No. **2**, 252–259.

[16] McKee S.P. (1981) A local mechanism for differential velocity detection, *Vision Research* **21**, 491–500.

[17] McKee S.P. (1983) The spatial requirements for fine stereoacuity, *Vision Research* **23**, 191–198.

[18] Nakayama K. (1985) Higher order derivatives of the optical velocity vector field: limitations imposed by biological hardware, In *Brain mechanisms and spatial vision.* (Edited by Igle, D., Jeannerod, M. and Lee, D.). Martinus Nijhoff, Holland, 1985.

[19] Nakayama K. (1985) Biological image motion processing: a review, *Vision Research* **25**, 625–660.

[20] Nakayama K. and Tyler C.W. (1981) Psychophysical isolation of movement sensitivity by removal of familiar position cues, *Vision Research* **21**, 427–433.

[21] Rashbass C. (1970) The visibility of transient changes of luminance, *Journal of Physiology* **210**, 165-186.

[22] Rashbass C. (1970) Unification of two contrasting models of the visual incremental threshold, *Vision Research* **16**, 1281–1283.

[23] Regan D.M., Kaufman L. and Lincoln J. (1986) Motion in depth and visual acceleration, In *Handbook of Human Perception and Human Performance,* Vol. I, Chapter 19. (Edited by Boff, K.R., Kaufman, L. and Thomas, J.P.) John Wiley & Sons, Inc., 1986.

[24] Snowden R.J. and Braddick O.J. (1991) The temporal integration and resolution of velocity signals, *Vision Research* **31**, 907–914.

[25] Schwartz B.J. and kaufman L. (1987) Psychophysical measures for changes in direction of motion, *Society of Neuroscience Abstracts* **13**, 631.

[26] Watson A.B. (1987) Estimation of local spatial scale, *J. of the Optical Society America A* **4**, No. **8**, 1579–1582.

[27] Watson A.B., Ahumada A.J. and Farrell J.E. (1986) Window of visibility: a psychophysical theory of fidelity in time-sampled visual motion displays, *J. of the Optical Society America A* **3**, No. **3**, 300–307.

[28] Watt R.J. (1985) Image segmentation at contour intersections in human focal vision, *J. of the Optical Society America A* **2**, No. **7**, 1200–1204.

[29] Werkhoven P. and Koenderink J.J. (1991) Visual processing of rotary motion, *Perception & Psychophysics* **49**, No. **1**, 73-82.

[30] Wertheim T. (1894) Uber die indirekte Sehschärfe, *Z. Psych. Physiology Sinnesorg* **7**, 172–189.

[31] Westheimer G. (1982) The spatial grain of the perifoveal visual field, *Vision Research* **22**, 157–162.

Figure captions

Figure 1: A sketch of some modulation functions: pulse shaped $\delta(t)$, triangular $\Lambda(t)$, sinusoidal $\Omega(t)$, block shaped $\Pi(t)$ and sample function $\Upsilon(t)$, as a function of time t. For this illustration, all functions are normalized such that their mean value over time is 0.5, their temporal wavelength is 2π, and the modulation amplitude is 100%, except for function $\Upsilon(t)$. The function $\Upsilon(t)$ is the velocity modulation function as used in the 'sample and hold' paradigm of Burr *et al.* (1986) (for this illustration also with mean 0.5, and temporal wavelength 2π).

Figure 2: Threshold speed modulation amplitudes W_v as a function of speed modulation frequency ω. Thresholds W_v are the relative speed modulation amplitudes (dv_x/v_0) that yield 80% correct answers. The (very similar) data of three observers (HS, PW and AT) have been averaged. Triangular symbols indicate (symmetric) triangular speed modulation functions $\Lambda(t)$. The different sizes of the symbols indicate different average speeds v_0 as given in the figure. Note, that the closed triangles indicate a special condition in which the diffusing screen was removed such that the target was a luminous dot. Square symbols indicate thresholds for (symmetric) block speed modulation functions $\Pi(t)$.
The + symbols indicate results obtained in a separate speed discrimination experiment. Observers indicated whether a uniform motion stimulus moved at a high velocity ($v_0 + dv_x$) or at a low velocity ($v_0 - dv_x$). As in the modulation experiments, $W_v = dv_x/v_0$. For the left speed discrimination threshold in the figure the presentation time of each speed interval was 1 s, for the right threshold it was 125 ms. To facilitate a comparison with the speed *modulation* thresholds, the two speed *discrimination* thresholds are plotted at a horizontal position that equals half their inverse presentation time.
Half the length of the shown error bar for each data point corresponds to the square-root-variance of the binomial probability distribution for that point.

Figure 3: Upper cut-off modulation frequency ω_c as a function of speed v_0. Cut-off frequency ω_c is defined as the modulation frequency yielding threshold performance (80% correct answers) for a modulation amplitude of 100%. Cut-off frequencies are measured for triangular speed modulation function $\Lambda(t)$ and two individual observers (HS and AT). Parameter settings are listed in Table 2. Closed symbols indicate cut-off frequencies for a luminous target dot.
Open symbols are cut-off frequencies for a blurred (blob shaped) target and are extrapolated from the data of Fig. 2 using the temporal low-pass filter described in the Model Section. Because these extrapolated data for observers AT and HS were very similar, we averaged them for this presentation.

Figure 4: The dependence of upper cut-off frequency ω_c on speed v_0 for a triangular speed modulation function $\Lambda(t)$ with fixed 80% speed modulation amplitude. Eccentricity was *fixed* at 10 deg. Individual data for two observers (AT and HS) are plotted.

Figure 5: Threshold direction modulation detection amplitudes $d\theta$ as a function of modulation frequency ω. Weber fractions W_d are simply related to $d\theta$ by the expression: $W_d = \tan d\theta$. Data of two observers (AT and HS) have been averaged. Filled circles are data obtained with a sinusoidal motion direction modulation function $\Omega(t)$; Open squares indicate a block shaped motion direction modulation function $\Pi(t)$ (corresponding to a vertical position modulation $\Lambda(t)$).

Figure 6: Modulation thresholds for the detection of pulse shaped direction modulation functions $\delta(t)$, expressed as vertical *spatial excursions* of the block shaped vertical position modulation $\Pi(t)$. Spatial thresholds are shown as a function of modulation frequency ω. The average speed v_0 was 1.7 deg/s and the eccentricity of presentation was 3.4 deg. Data for two observers (AT and HS) have been averaged.

Figure 7: **Probabilities P of correct answers as a function of the *normalized* velocity vector modulation amplitude ξ. With a normalized vector modulation amplitude ξ, we mean a modulation amplitude for a given parameter setting, divided by the modulation amplitude yielding threshold performance for that particular setting.**
The data points are collected using all speed/direction modulation detection experiments described in this paper. The horizontal axis represents normalized modulation amplitudes ξ. The ordinate represents the percentages correct P for a small range of normalized modulation amplitudes clustered around a range of plotted normalized amplitudes ξ_i of the data points (filled circles). Half the length of the plotted error bar for a normalized amplitude ξ_i corresponds to the square-root-variance or standard deviation (σ_i) for that point (see text).
Because we normalized the amplitudes to the threshold amplitude, the curve is expected to reach threshold (80% correct answers) for $\xi = 1$ and reach chance level (50%) for $\xi = 0$. The solid curve is a best fit of the psychometric function $Erf(z)$ to the data, taking the squared modulation amplitude $z = \xi^2$ as its argument. The dotted and dashed curves are the expected psychometric curves for arguments $z = \xi$ and $z = \xi^3$ respectively, and fit less well.

Figure 8: Two similar (but fundamentally different) modulation detection processing streams. The first stream (a) consists of the extraction of a *true velocity signal*, followed by low-pass temporal filtering and by a final detection stage (*e.g.*, peak or variance detection). The second stream (b) consists of a velocity extraction that does not yield a true velocity signal but a *low-pass transformation of the true velocity*, followed by a final detection stage as in the first stream.

Figure 9: A standard motion detector (Reichardt correlator). Standard motion analysis consists of two input lines (receptive fields) with temporal filters $f(t)$, a delay filter (with time constant τ) for one of the input lines, a correlation stage and a temporal integration filter $\int_T$.

Figure 10: Ensemble activation profiles as a function of time. The ensemble of motion detectors considered in this figure is parameterized by a single parameter, the tuning velocity v_t (vertical axis). Activation profiles are given as a function time t (horizontal axis) for a triangular velocity modulation function (upper left corner) and a constant speed (lower left corner). At the right side we show the resulting activation profiles for both functions after temporal integration ($\int_T$) of each motion detector output in the ensemble (*i.e.*, integration along the horizontal axis).

Tables

wave form	speed v_0 (deg/s)	sweep-length d_0 (cm)	sweep-time t_0 (s)	distance d_v (m)	eccentricity ϵ (deg)
$\Lambda(t)$	1.0	43	4	6.00	4.10
$\Lambda(t)$	1.7	30	1	10.20	1.68
$\Lambda(t)$	5.0	30	1	3.40	5.04
$\Lambda(t)$	15.0	30	1	1.12	15.0
$\Pi(t)$	2.5	30	2	3.40	5.04

Table 1: *Parameter settings for speed modulation thresholds of Fig. 2.*

wave form	speed v_0 (deg/s)	sweep-length d_0 (cm)	sweep-time t_0 (s)	distance d_v (m)	eccentricity ϵ (deg)
$\Lambda(t)$	0.26	23	2	25.0	0.52
$\Lambda(t)$	0.63	23	2	10.4	1.26
$\Lambda(t)$	2.5	23	2	2.60	5.00
$\Lambda(t)$	5.0	35	2	2.00	10.0
$\Lambda(t)$	7.5	35	2	1.35	15.0

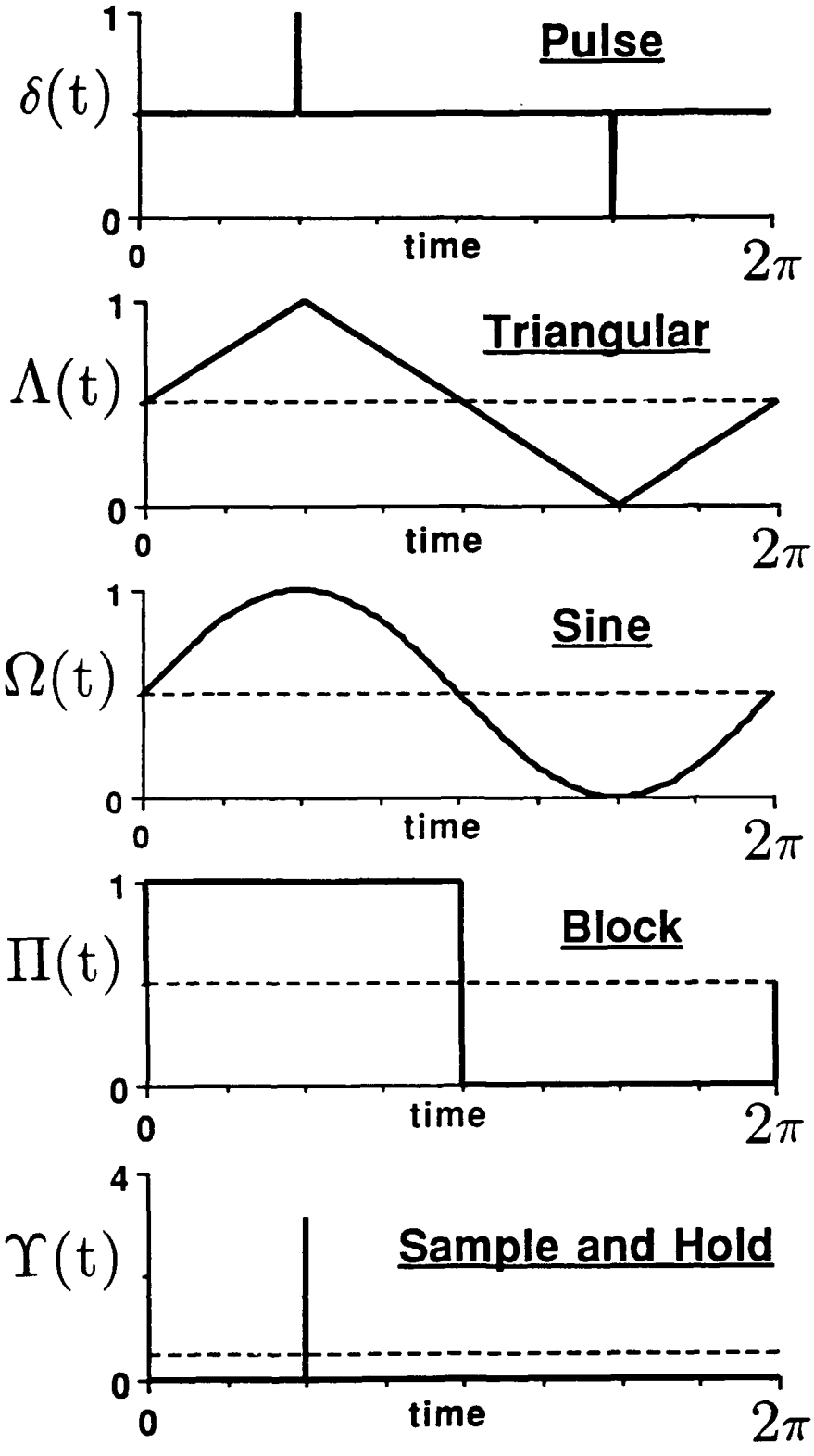
Table 2: *Parameter settings for speed modulation thresholds of Fig. 3.*

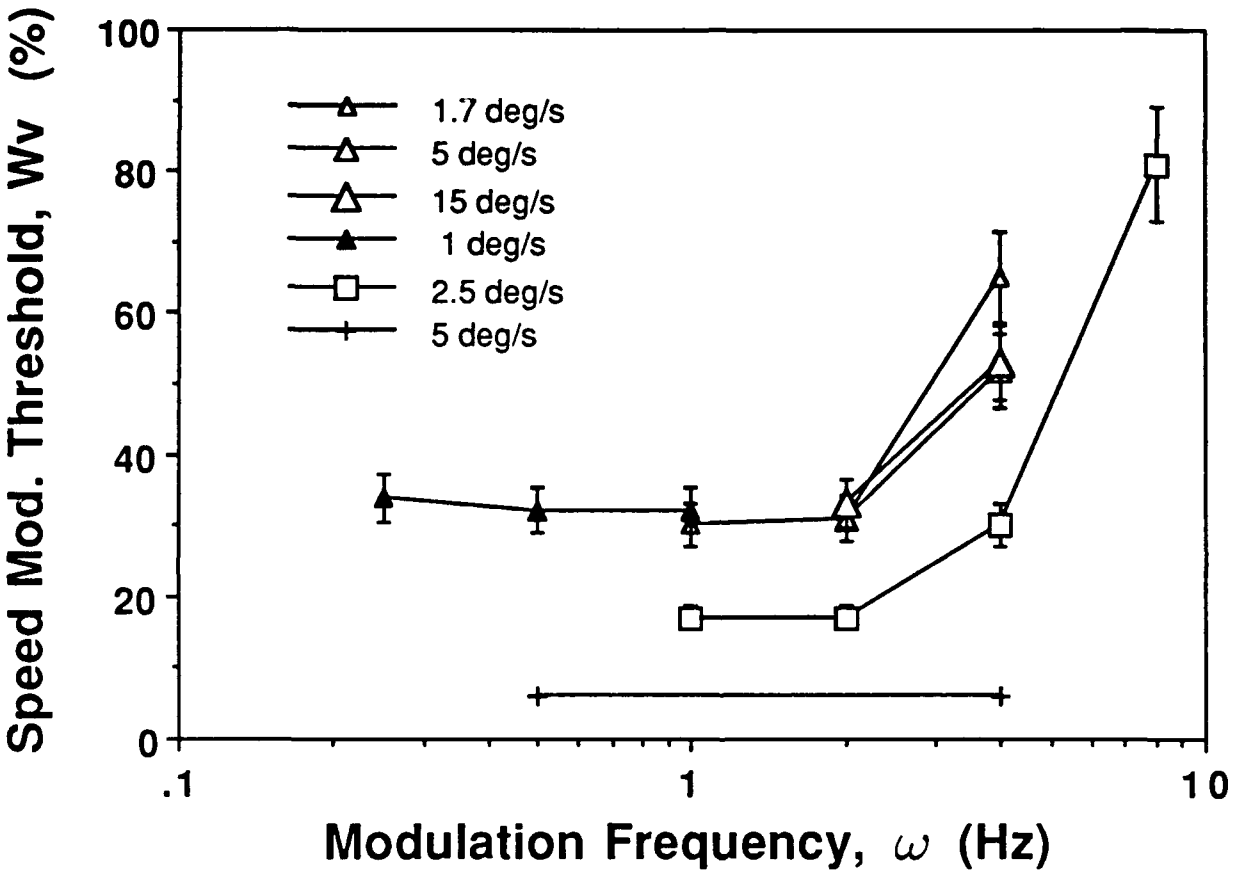
eccentricity ϵ (deg)	cut-off HS ω_c (Hz)	cut-off AT ω_c (Hz)	cut-off PW ω_c (Hz)
0.5	n.a.	6.0	6.8
5	5.3	5.5	5.4
10	5.2	5.3	4.9
15	n.a.	5.7	5.8

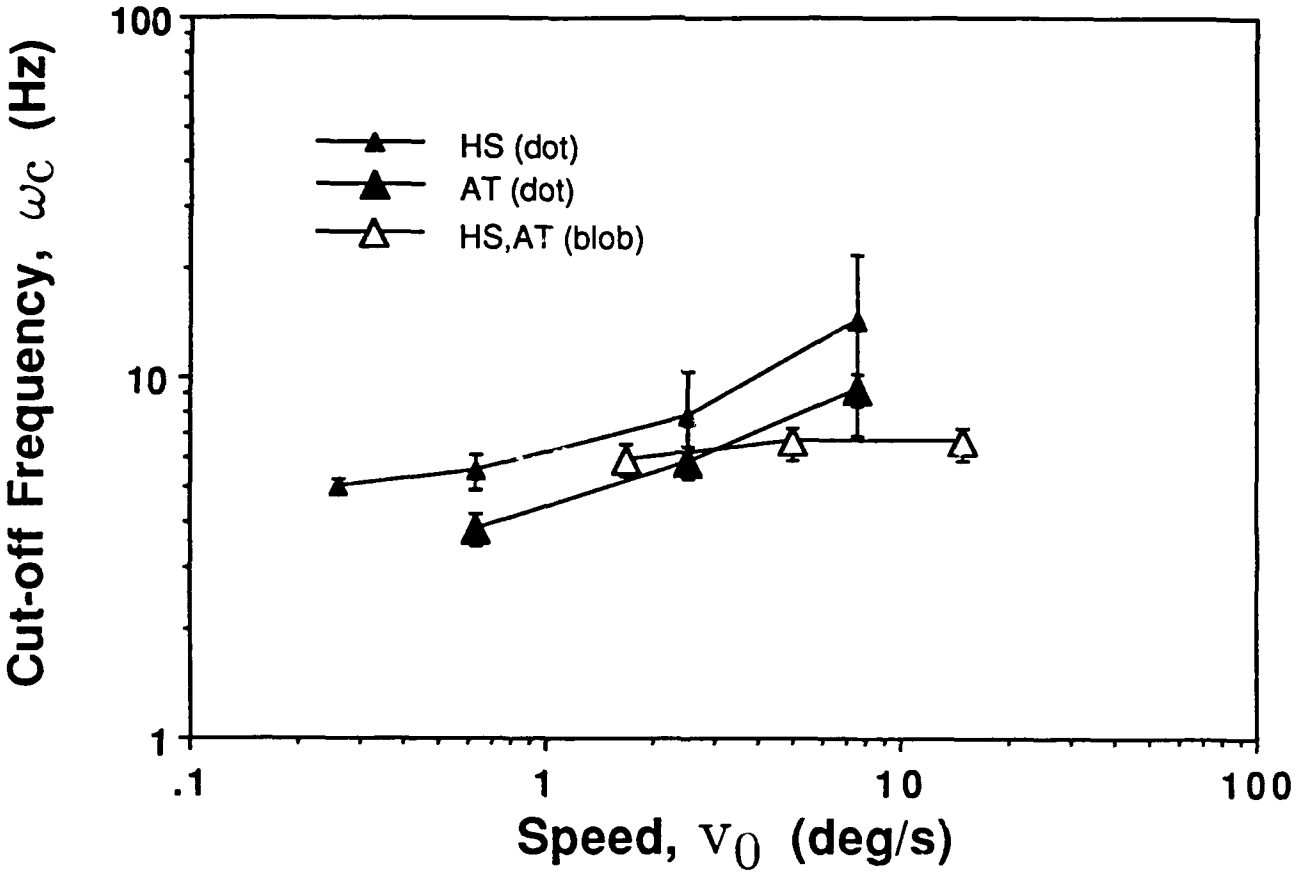
Table 3: *Cut-off frequency ω_c obtained for the triangular speed modulation function $\Lambda(t)$, with relative amplitude 80% and mean velocity $v_0 = 4$ deg/s, for three observers at four eccentricities ϵ. The data for observer HS at 0.5 and 15 deg eccentricity were not available.*

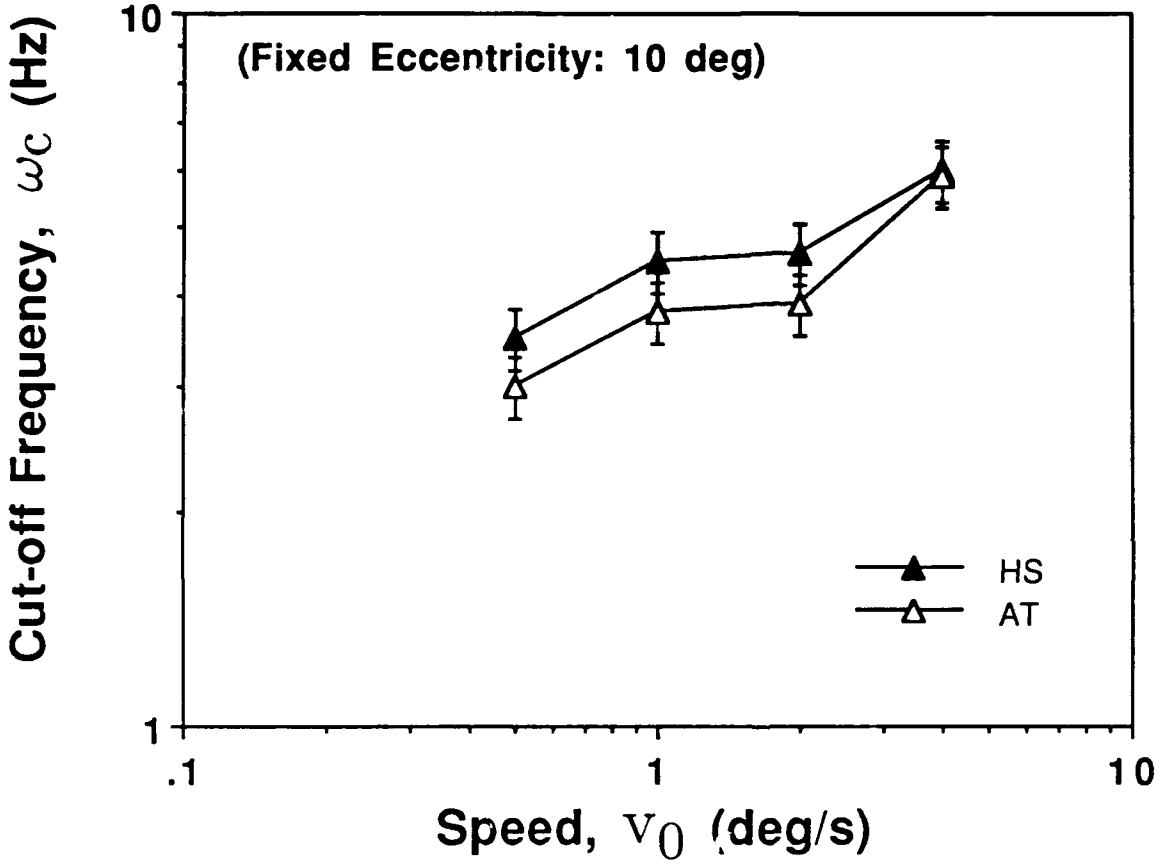
wave form	speed v_0 (deg/s)	sweep-length d_0 (cm)	sweep-time t_0 (s)	distance d_v (m)	eccentricity ϵ (deg)
$\Omega(t)$	1.0	43	4	6.00	4.10
$\Omega(t)$	2.0	43	2	6.00	4.10
$\Pi(t)$	1.7	36	2	6.00	3.43
$\delta(t)$	1.7	36	2	6.00	3.43

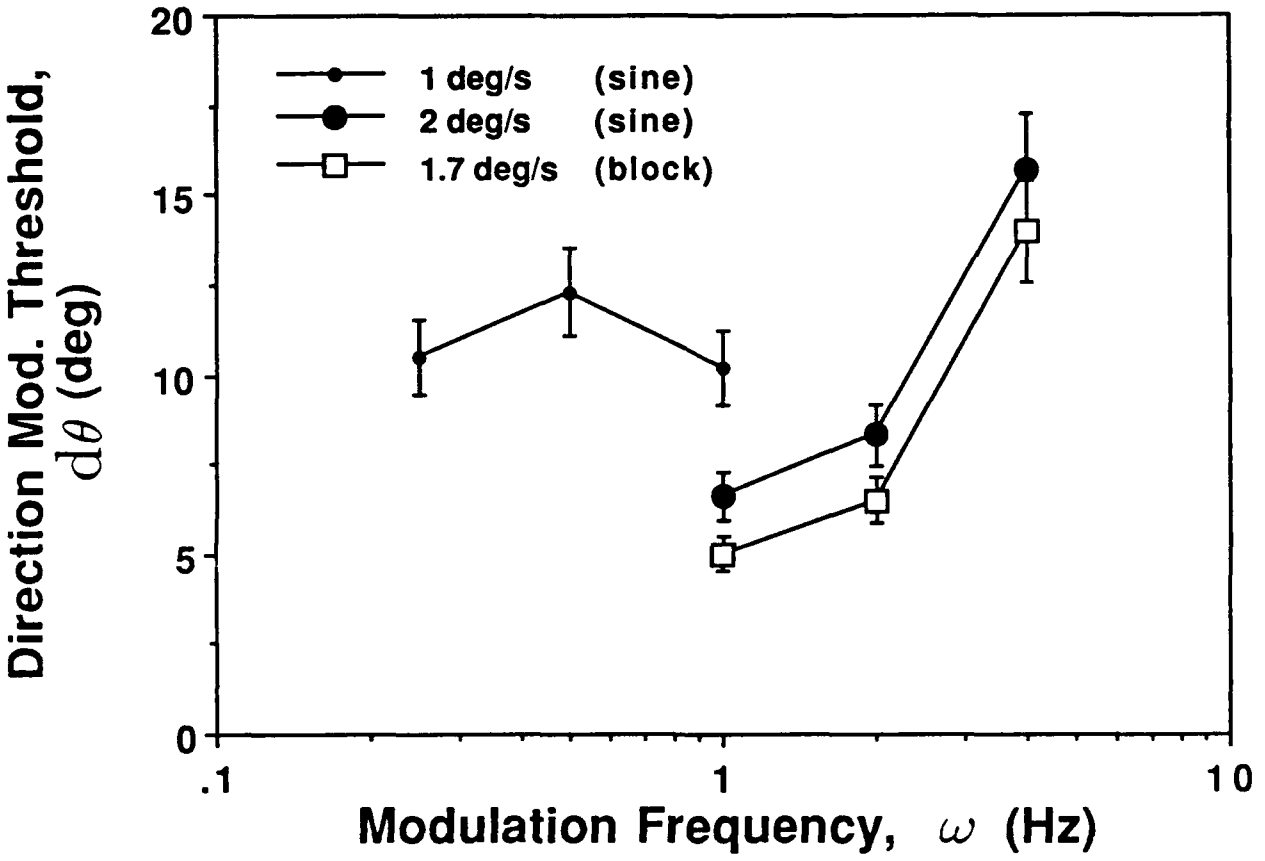
Table 4: *Parameter settings for direction modulation thresholds of Fig. 5 and Fig. 6.*

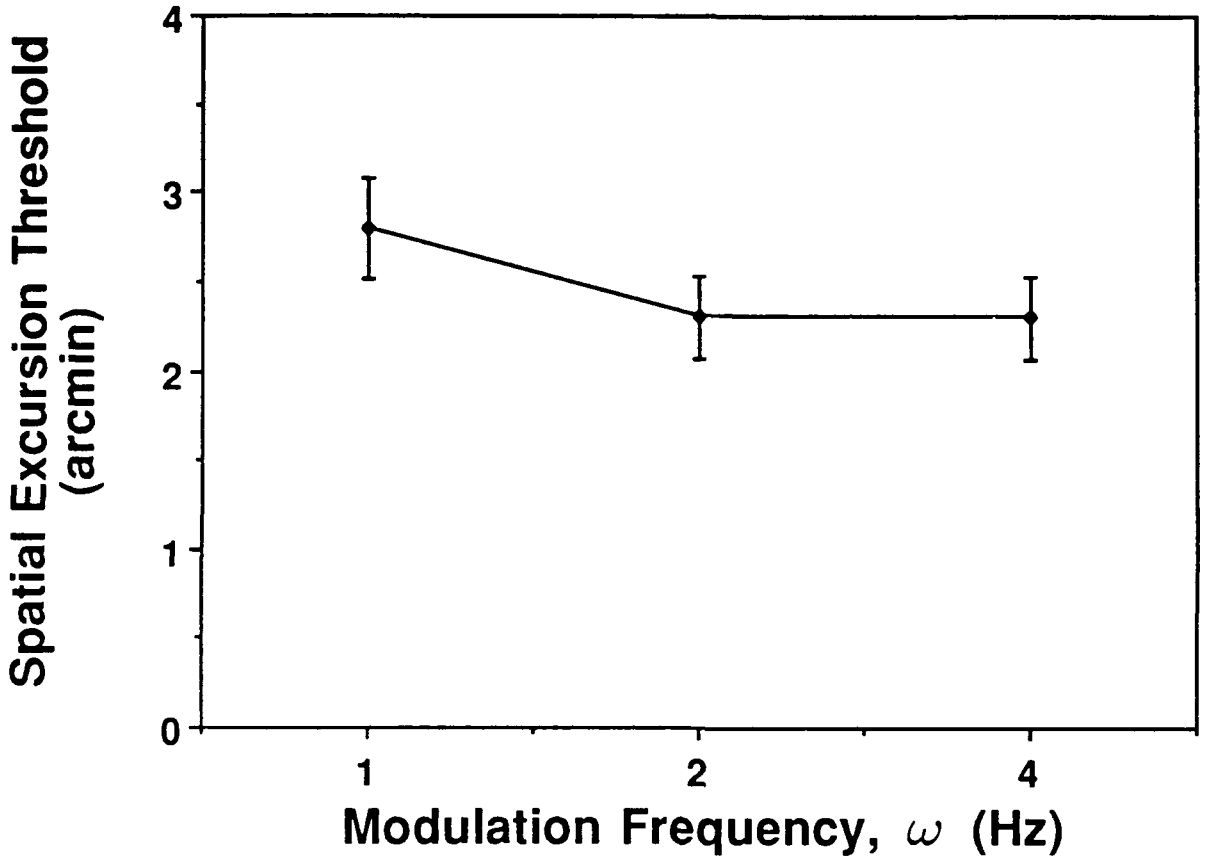

Pulse
δ(t)
1
0
0
time
2π
Triangular
Λ(t)
1
0
0
time
2π
Sine
Ω(t)
1
0
0
time
2π
Block
Π(t)
1
0
0
time
2π
Sample and Hold
ϒ(t)
4
0
0
time
2π

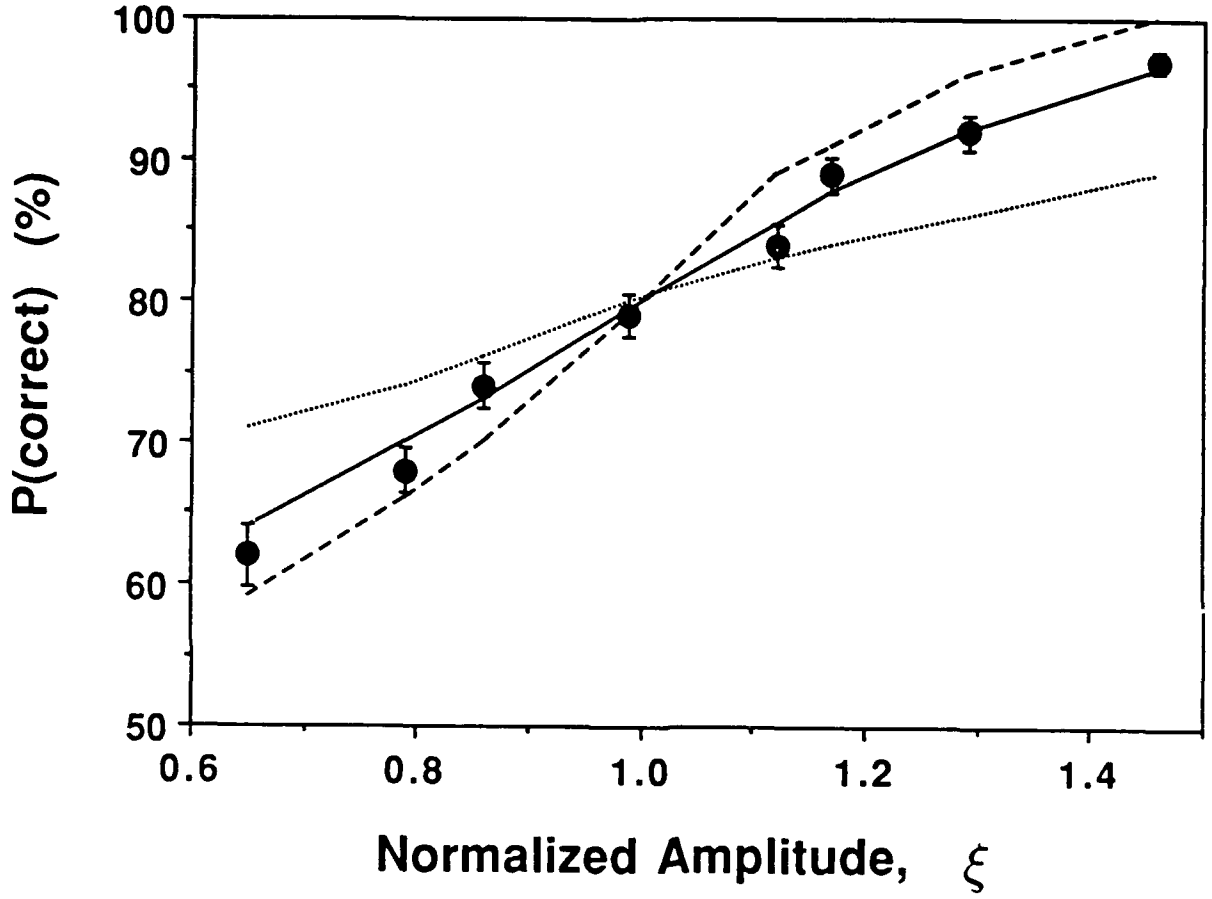

Speed Mod. Threshold, Wv (%)
100
80
60
40
20
0
.1
1
10
Modulation Frequency, ω (Hz)
1.7 deg/s
5 deg/s
15 deg/s
1 deg/s
2.5 deg/s
5 deg/s

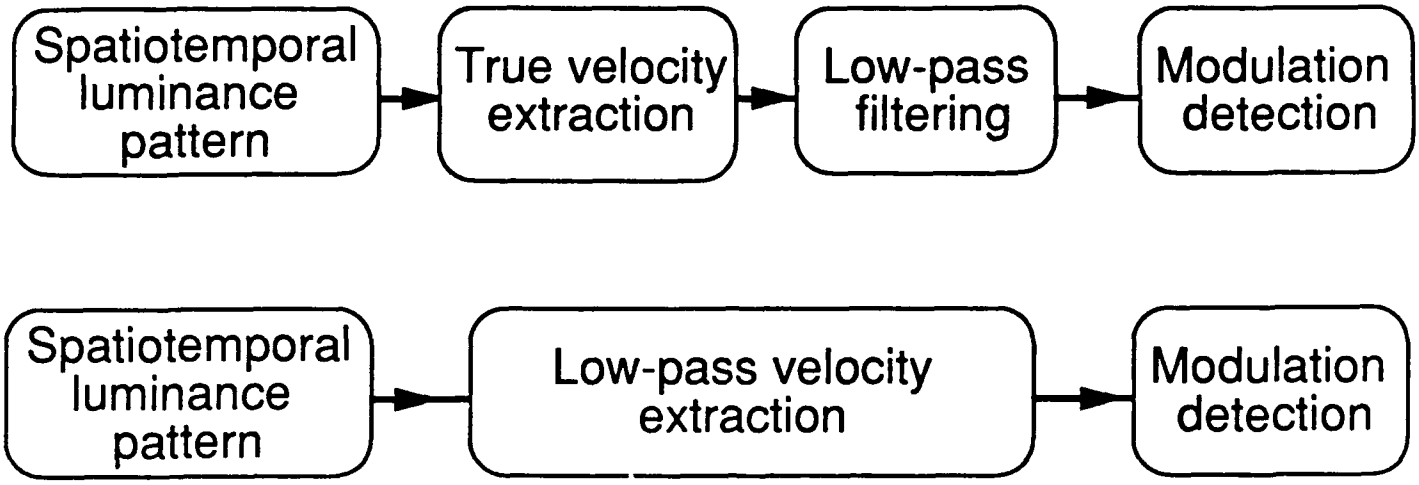

100
10
1
.1
1
10
100
HS (dot)
AT (dot)
HS,AT (blob)
Cut-off Frequency, ω_c (Hz)
Speed, v_0 (deg/s)

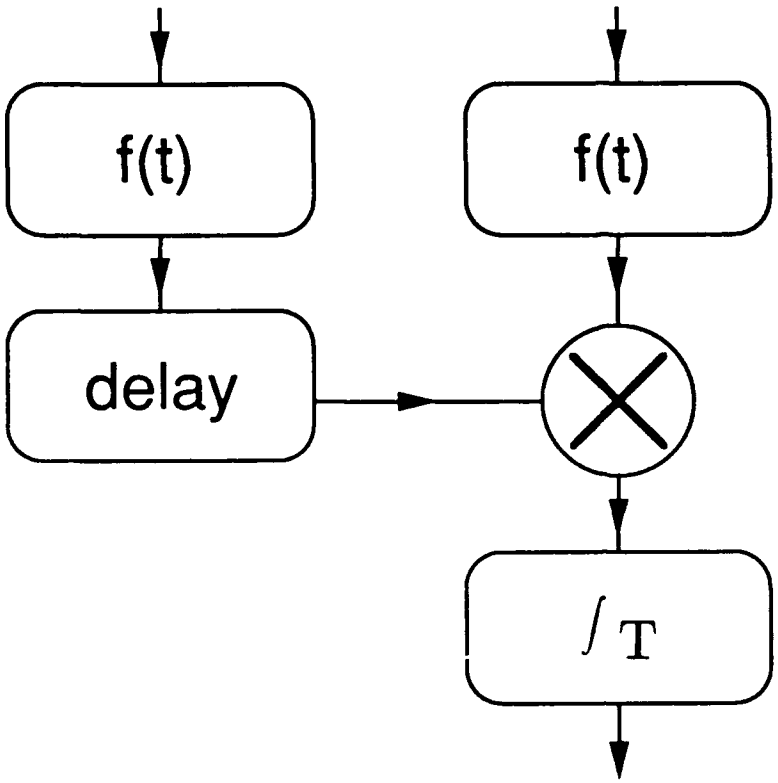

(Fixed Eccentricity: 10 deg)
Cut-off Frequency, ω_c (Hz)
10
1
.1
1
10
Speed, v_0 (deg/s)
HS
AT


20
15
10
5
0
.1
1
10
1 deg/s (sine)
2 deg/s (sine)
1.7 deg/s (block)
Direction Mod. Threshold, dθ (deg)
Modulation Frequency, ω (Hz)


Spatial Excursion Threshold
(arcmin)
4
3
2
1
0
1
2
4
Modulation Frequency, ω (Hz)

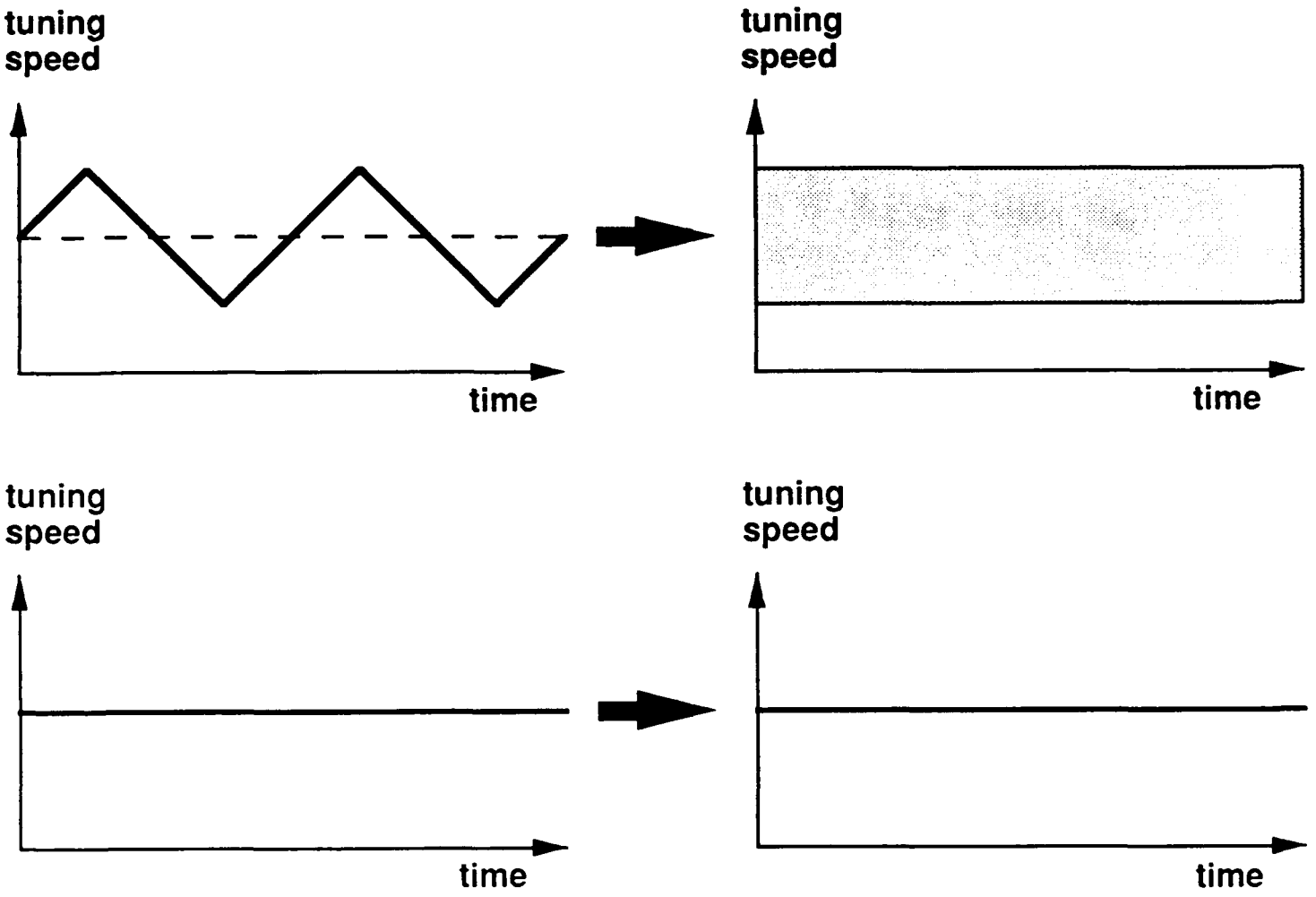

100
90
80
70
60
50
P(correct) (%)
0.6
0.8
1.0
1.2
1.4
Normalized Amplitude, ξ


Spatiotemporal luminance pattern
True velocity extraction
Low-pass filtering
Modulation detection
Spatiotemporal luminance pattern
Low-pass velocity extraction
Modulation detection


f(t)
f(t)
delay
$\int_T$

Before Filtering

After Filtering



Pulse Modulation Detection in Human Motion Vision

Herman P. Snippe[1] and Peter Werkhoven[2]

Department of Psychology and Center of Neural Science,
New York University,
6 Washington Place, New York 10003

March 10, 1992

(PREPRINT — Subm. for publ. Vision Research, Feb 27 1992)

[1] On leave from the Utrecht Biophysics Institute, Utrecht University, P.O. 80000, 3508 TA Utrecht, The Netherlands.

[2] Future address (as of May 15, 1992): Utrecht Biophysics Institute, Utrecht University, P.O. 80000, 3508 TA Utrecht, The Netherlands.

Abstract

We present data on the human sensitivity to temporal pulse modulations of target velocity. We measured threshold detection modulation amplitudes for pulse-shaped speed modulations, as a function of pulse duration and temporal frequency.

At short pulse durations (up to 50 ms) and low modulation frequency (1 Hz), detection amplitudes are ruled by a Bloch law: the product of pulse duration and threshold modulation amplitude is a constant. This constant corresponds to a position modulation with an amplitude of 3 arcmin in a coordinate frame that moves at the average speed (3 deg/s) of the target. At longer pulse durations we find deviations from Bloch's law. Speed modulation thresholds are not critically dependent on target luminance contrast.

These results are modeled by a modulation detection process in two stages. A functional description of the first stage is filtering of the true speed modulation signal by a second order low-pass filter with a characteristic time constant of 20-25 ms. The second (decision) stage is variance detection: modulations are detected when the variance of the filtered modulation function exceeds a certain threshold variance. The square root threshold variance is estimated 8-10%.

This two-parameter model accurately predicts the measured dependence of pulse modulation detection thresholds on pulse duration and pulse density.

1 Introduction

Structure-from-Motion theories (Arnspang, 1988; Koenderink, 1986) show that the spatiotemporal dynamics of motion fields contain useful information for building representations of our 3-D environment. An intriguing question is how sensitive man is to modulations in motion parameters (*e.g.* speed and direction) and what mechanisms have evolved for extracting these modulations.

This question motivated Werkhoven *et al.* (1992) to perform experiments on the detection of temporal modulations of speed and direction (optic acceleration). They have presented evidence that the human visual system does not detect these modulations based on a optic acceleration signal. Modulation detection is accurately modeled by a variance detection process based on a low-pass transformation of the true velocity signal.

The above conclusions were based on modulation transfer functions: the dependence of detection thresholds on modulation frequency. Transfer functions are powerful predictors for the performance of linear systems and have a long history. For example, transfer functions revealed the characteristics of luminance modulation detection in the human visual system (de Lange, 1958).

An alternative tool for studying human detection performance in visual research has an even longer history: pulse detection. For example, pulse detection experiments lead to the formulation of Bloch's law for luminance pulse detection (Bloch, 1885). Pulse amplitudes yielding threshold detection performance, as a function of pulse duration, can reveal the linearity and temporal characteristics of the visual detection mechanism. Furthermore, a study of detection

thresholds as a function of pulse *density* can yield information on the type of decision process that rules the detection process (Roufs, 1974; Watson, 1979).

The conclusions inferred from pulse detection experiments and those from modulation transfer functions, have been shown to be consistent for luminance modulation detection experiments (Roufs, 1972; Roufs, 1973). This consistency proves the linearity of the temporal filter stage that rules the detection of temporal luminance modulations. Furthermore, consistency provides evidence that modulation transfer functions are determined by a *single* temporal filter (Koenderink, 1978).

In this paper, we present a study on the detection of *pulse-shaped* speed modulation functions and examine the consistency of the present findings with the modulation transfer functions and model presented in Werkhoven *et al.* (1992).

2 Methods

In short, stimuli consisted of dot targets that moved along horizontal trajectories at a speed either modulated in time or kept constant. Observers indicated whether the target motion as constant or modulated.

2.1 Apparatus

We used an Atari 1040 ST computer to generate the stimuli and to control the experimental procedures. An Atari SM125 high resolution 70Hz white phosphor monochrome monitor was used (luminance 71 cd/m^2) to display the stimuli. The phosphor decay rate was approximately 0.3 msec. The spatial resolution of the SM125 was 640 × 400 pixels. Viewing distance was constant throughout the experiment (210 cm), yielding angular screen dimensions of 5.6 × 3.5 deg.

2.2 Stimulus

A 'snapshot' of the spatial configuration of the targets is shown in Fig. 1.

— *Figure 1 about here* —

The stimulus consists of two rows of four targets (squares of 8 arcmin width). The horizontal separation between targets in a row was 1.3 deg. In the first row, targets moved coherently at a *constant* speed v_0 (uniform motion path) in a *random horizontal* direction. In the second row (at the other side of the fixation mark), the targets moved coherently with a *modulated* speed $v(t) = \delta(t)$ (modulated motion path) in the *opposite* direction of the first row. The vertical position of the first row (top or bottom) was randomized. The opposite motion of the

constant and modulated path discouraged observers to use the relative horizontal *positions* of the targets in the two paths as a cue for modulation detection.

The vertical separation of the top and bottom row was 3.2 deg. The traversed target trajectory length was 5.2 deg. Whenever a target reached the end of a trajectory, it returned to the start of its trajectory and continued its motion. As a result, at every instant of time, four targets were visible in each row. Observers fixated their view on the fixation mark (see Fig. 1) which was placed half way of the vertical positions of the two rows and half way of the trajectory. Consequently, the eccentricity of both motion paths varied between 1.6 and 3.1 deg. Targets were presented dark on a light background so that the visual system of the observers adapted to a well-defined photopic luminance level.

The average speed of the traversing targets was $v_0 = 3.0$ deg/s (or 5 pixels per frame) in all experiments reported in this paper. The transition time (the time to cross the trajectory) was 1.71 s (120 frames). Total presentation time was fixed at 4.1 seconds. The targets appeared instantaneously at the start of their trajectory and, likewise, disappeared instantaneously at the end of their trajectory. The positional phase (the initial position of the targets in a row) and the phase of the speed modulation function were both randomized between presentations.

2.3 Pulse Modulation Functions

The top and bottom row moved rigidly in opposite directions with a velocity modulation function $v(t) = \delta(t)$. We used a three parameter pulse modulation function $\delta(t)$ as shown in Fig. 2. Basically, the modulation function $\delta(t)$ consists of alternating *positive* and *negative* pulses (with equal magnitude dv and duration α) superimposed on a constant speed v_0.

— *Figure 2 about here* —

The three parameters that determine the modulation function $\delta(t)$ are:

1. Pulse duration α. The values for α were limited to integer multiples of the inverse monitor refresh rate (14.3 ms).

2. Pulse modulation amplitude dv/v_0. Because the average speed was fixed at 5 pixels/frame and because the values of dv were limited to an integer number of pixels per frame, the modulation amplitudes dv/v_0 were 0%, 20%, 40%, 60%, 80% or 100%. We did not use modulation amplitudes higher that 100% yielding local reversals of motion direction for the negative velocity pulses.

3. The temporal separation β between the onset of consecutive pulses. Because the sign of the amplitudes of the pulses alternated in time, the average speed was constant (v_0), independent of modulation amplitude dv. Thus, the average speed of the uniform and modulated path are equal.

2.4 Psychophysical Procedure

Speed modulation thresholds were measured in a modulation *detection* experiment. In one session, observers viewed 25 stimulus presentations with a constant average speed v_0, modulation amplitude dv, pulse duration α and pulse separation β. The modulated motion path appeared either at the top vertical position or in the bottom position (its position was randomized). The uniform motion path always appeared in the position opposite to the modulated path. The phase of the modulation function was randomized. The task of the observers was to indicate the position (top/bottom) of the modulated motion path by means of a joy-stick.

Usually 4-5 sessions with different adequately chosen modulation amplitudes were sufficient to determine the speed modulation detection threshold (for a given constant pulse duration and pulse separation) by data interpolation. We defined the speed modulation threshold W as the relative modulation amplitude dv/v_0 at threshold performance (yielding 80% correct answers). Measurements were performed binocularly with natural pupils in a darkened room. No feedback was provided in either experiment.

2.5 Observers

The two authors (HS and PW), both experienced observers in visual motion experiments, were subjects throughout the experiment. The general findings were confirmed by a third, naive subject.

3 Experiments

3.1 Pulse duration

3.1.1 Results

We collected data on the dependence of speed modulation thresholds W on pulse duration α for a fixed temporal interval $\beta = 500$ ms between the onset of consecutive pulses (see Fig. 3[1]).

— Figure 3 about here —

Detection thresholds are 120% (HS) and 140% (PW) for the shortest pulse duration (14.3 ms) and decrease monotonically with increasing pulse duration, down to 20% for a long pulse duration of 143 ms (HS) and 158 ms (PW).

[1] It should be noted that the rightmost data point in Fig. 3 was determined using the reverse of the procedure described in the methods section, that is, for this data point we fixed the modulation amplitude at 20% and determined the pulse duration α yielding threshold performance. The leftmost data point (at α=14.3 ms) is extrapolated from the probabilities correct at modulation amplitudes smaller than 100%. Modulation amplitudes higher than 100% yield a reversal of motion direction and were not used.

The guide lines are theoretical relations derived in the Model Section, and show the predicted asymptotic behaviour for small pulse durations (the solid line) and large pulse durations (dashed line).

3.1.2 Discussion: Small Pulse Durations and Blochs's Law

Figure 3 shows a dependence of pulse modulation thresholds on pulse duration that asymptotes to a line with a slope of approximately -1 for the shortest pulse durations examined in log-log coordinates. This indicates that the product of modulation W and pulse duration α is constant at threshold (Bloch's law). The product $W\alpha$ is exactly the integral of the modulation pulse, that is, the spatial displacement of the dot in a coordinate frame that moves at speed v_0. This suggests that, for short pulse durations, modulation detection is based on a spatial cue: the spatial excursion of the dot in the moving frame. The threshold spatial excursion is $Wv_0\alpha$ and is approximately 3 arcmin for observer HS and 3.5 arcmin for PW at the average speed of 3 deg/s in this experiment.

Spatial excursion is the temporal integral of the velocity modulation function. Thus, the finding that the magnitude of the spatial excursion rules detection performance strongly suggests that the detection stage is based on the integral of the velocity signal. A likely implementation of velocity integration is a *low-pass temporal filter* operating on the true velocity signal. The time constant of this low-pass filter determines the integration time and thus the range of pulse durations for which we may expect a Bloch law. In the Model Section, we will elaborate on this issue, and show that this finding is consistent with the model proposed by Werkhoven *et al.* (1992) for the detection of different types modulation functions in speed and direction.

3.1.3 Discussion: Large Pulse Duration and Low-pass Filtering

For larger pulse durations (*e.g.*, α = 143 ms) pulse modulation thresholds deviate from the asymptote behavior described by a Bloch law.

In terms of a low-pass filter stage, this performance at large pulse duration is expected when the pulse duration exceeds the time constant τ of the temporal filter. Now, the filter response is no longer the impulse response function (as for short pulse durations), but can be described by a response to an incremental step (at the on flank of the pulse) followed by a response to a decremental step (at the off flank). For large pulse durations ($\alpha \gg \tau$), asymptotic threshold performance becomes independent of the time constant and depends exclusively on the pulse duration. The deviations from a Bloch law reveal information about the characteristic time constant of the low-pass filter (see the Model Section).

3.2 Pulse Separation

In the above experiment, we measured pulse modulation thresholds as a function of pulse duration α at fixed pulse separation β = 500 ms. Here, we study modulation thresholds as a

function of pulse separation β for pulse durations $\alpha = 14.3, 28.6$ and 42.9 ms.

3.2.1 Motivation and Predictions

Roufs (1974) performed modulation detection experiments in the *luminance* domain and showed that detection thresholds decrease with the number of pulses in the stimulus presentation. This decrease was in agreement with the steepness of the psychometric functions obtained in his experiments. We may expect similar behavior in the *motion* domain. that is, detection thresholds decrease with increasing pulse density $(1/\beta)$.

However, when the pulse density reaches a point where the pulse separation β is comparable to the time constant τ of a hypothesized low-pass filter, that operates on the modulation function, the pulse responses of the temporal filter start to interfere. Now, the temporal filter integration of the the modulation function yields an annihilation of neighboring positive and negative pulses. This would lead to an increases in thresholds for pulse separations smaller than the characteristic time constant τ of the low-pass filter.

To summarize, modulation thresholds are expected to decrease as a function of modulation frequency (due to increased pulse rates) until the pulse separation is of the order of the time constant τ. For even smaller separations, thresholds will increase with frequency (due to temporal integration).

3.2.2 Results

We measured modulation thresholds as a function of pulse separation β for short pulse durations $\alpha_1 = 14.3$ ms, $\alpha_2 = 28.6$ ms and $\alpha_3 = 42.9$ ms (in the range of Bloch's Law). As mentioned in the Method section, the number of available modulation amplitude levels is limited. Therefore, we used the percentages correct answers at pulse durations α_1, α_2 and α_3 and all available modulation amplitudes to estimate a single threshold W for α_1. The argument runs as follows. As we have seen in Fig. 3, the threshold modulation amplitudes W decrease with increasing pulse duration (Bloch's law) for *small* pulse durations. However, the *product* of modulation amplitude W and pulse duration α (pulse area, Ω) at threshold performance is constant in this range. Hence, we can use the thresholds at α_2 and at α_3 to estimate the threshold at α_1. Therefore, we computed the pulse areas for these three pulse durations and their corresponding percentages correct answers for a given pulse separation β, yielding a psychometric function: percentages correct answers as a function of pulse area Ω. From this function, we estimated threshold pulse area Ω_0 and modulation threshold $W = \Omega_0/\alpha_1$ for pulse duration α_1.

— Figure 4 about here —

Figure 4 shows estimated thresholds W at pulse duration $\alpha_1 = 14.3$ ms as a function of the fundamental frequency ω of the modulation function. Note that modulation frequency ω

is inversely proportional to pulse separation: $\omega = 1/(2\beta)$. Thresholds first decrease from 120% (HS) and 140% (PW) to 54% (HS) and 78% (PW) with increasing modulation frequency up to 6 Hz. For frequencies larger than 6 Hz, thresholds increase with increasing frequency.

3.2.3 Discussion

Fig. 4 shows the predicted threshold dependence on pulse separation (or modulation frequency).

At low modulation frequencies, the pulse separation β is large compared to the time constant τ of the hypothesized temporal integration filter. For this condition the filter responses to each individual pulse are well separated in time and do *not* interfere. Now, thresholds are determined by the pulse density. Performance is expected to improve with pulse density assuming a variance detection stage or probability summation (see Model Section). Thus, thresholds decrease with increasing modulation frequency in the low frequency range.

At high modulation frequencies, the pulse separation becomes small compared to the time constant of the hypothesized low-pass filter and the filter response functions to individual pulses strongly interfere. Positive and negative pulses start to annihilate and performance deteriorates. For this condition, we consider the transfer function of a low-pass filter. The amplitudes of high frequent spectral components are reduced by low-pass filtering. Therefore, we expect the fundamental component (first order harmonic) to dominate the detection stage. In the Model Section, we derive the characteristic time constant τ of the hypothesized low-pass filter based on the above discussed dependence of detection thresholds on modulation frequency.

3.3 Luminance Contrast

3.3.1 Motivation

In order to evaluate 'Window of Visibility' theories applied to modulation detection (see General Discussion) it is useful to examine the dependence of human sensitivity to speed modulations for different luminance conditions. In the above experiments, the moving targets were drawn dark on a light background (a 100% contrast). Here, we measure threshold modulation amplitudes as a function of luminance contrast for two different modulation functions.

3.3.2 Results

One function had $\beta = 500$ ms and $\alpha = 28.6$ ms. The other function had $\beta = 56$ ms and $\alpha = 14.3$ ms. Luminance contrast was either 25%, 50% or 100% and was realised by drawing a corresponding percentage of the pixels of a dot dark (at random positions). The remaining pixels had background value.

— Figure 5 about here —

Results for two observers, and the two velocity pulse functions are shown in Fig. 5. Thresholds decrease slightly with increasing luminance contrast.

To further examine the effects of spatial structure of the dot target, observer HS (4D myopic) repeated the above experiments without his correcting spectacles at 100% luminance contrast. The measured thresholds without spectacles were nearly identical to those measured with spectacle correction.

3.3.3 Conclusion

The results show that modulation detection thresholds depend only slightly on luminance contrast and are independent of the amount of spatial blurring of the target. This suggests that spatial structure is not critical for speed modulation detection (see also Werkhoven *et al.*, 1992).

4 Model

Werkhoven *et al.* (1992) presented strong evidence for a two-stage speed modulation detection model. The first stage is a low-pass temporal transformation of the modulation function. The second stage is a variance detection process, based on the filtered modulation function. They showed that this two-stage detection model accurately describes the modulation transfer functions for the detection of differently shaped modulations (*e.g.*, triangular, block-shaped, sinusoidal) of target speed and direction.

In this section, we show that this model can also account for pulse modulations of speed.

.1 The first stage: A Second Order Low-pass Filter

The modulation function $\delta(t)$ is filtered by a second order low-pass temporal filter $g(t)$:

$$g(t) = \frac{t}{\tau^2}\, \epsilon^{-t/\tau}, \tag{1}$$

which has a transfer function $\tilde{g}(\omega)$ of temporal frequency $\omega = \frac{1}{2\beta}$:

$$\tilde{g}(\omega) = [1 + (2\pi\omega\tau)^2]^{-1}. \tag{2}$$

The resulting filtered modulation function is the convolution of this filter $g(t)$ with the modulation function $\delta(t)$: $g(t) \otimes \delta(t)$.

4.2 The Second Stage: Variance Detection

The psychometric functions presented by Werkhoven *et al.* (1992) clearly support the claim that observers based modulation detection on the *square* amplitude of the modulation function. Variance detection is an elegant and sufficient description of this phenomenon. It should be noted, however, that Rashbass (1976) has shown that a probability summation with summation coefficient 2 based on the *linear amplitude* of the (filtered) modulation function yields similar predictions.

We assume that the variance σ^2 of the filtered speed modulation signal has to exceed some fixed threshold value σ_0^2 for detection. The variance of the filtered modulation function is given by:

$$\sigma^2 = \frac{1}{2\beta}\int_{t=0}^{2\beta}[g(t) \otimes \delta(t) - v_0]^2 dt. \tag{3}$$

Note that the modulation functions, and thus their (linearly) filtered functions, are periodic in time with period 2β. In general, the expressions for the variance σ^2 are not elegant. However, we distinguish three conditions that allow simple expressions.

4.2.1 Non-interfering Pulses: $\alpha \ll \tau$

Here, we consider a condition where the pulse duration α of the modulation function $\delta(t)$ is very small compared to the time constant τ of filter $g(t)$ (ideal pulses). Furthermore, we take the pulse separation β such that the response function $g(t)$ to a given pulse becomes zero before the next pulse starts. For these conditions, the filtered modulation function is approximated by a series of non-interfering impulse response functions. The amplitude of this impulse response function is proportional to the pulse area αW. The variance $\sigma_p^2(\tau)$ of such impulse response function is simply:

$$\sigma_p^2(\tau) = \frac{1}{\beta}\int_{t=0}^{\beta}[W\alpha g(t)]^2 dt \approx \frac{W^2\alpha^2}{4\tau\beta} \qquad (\alpha \ll \tau). \tag{4}$$

The variance of non-interfering impulse response functions is inversely proportional to the characteristic filter constant τ.

4.2.2 Non-interfering Steps: $\tau \ll \alpha$

Consider a condition where the pulse duration α is finite and much larger than the time constant τ. Now, the modulation function $\delta(t)$ is described by two step functions: an incremental step W at time $t = 0$ and a decremental step $-W$ at time $t = \alpha$. A convolution of these step functions with the time filter $g(t)$ yields a variance σ_s^2:

$$\sigma_s^2(\tau) = \frac{1}{\beta}\int_{t=0}^{\beta}\left(W\int_{y=Max[t-\alpha,0]}^{t} g(y)dy\right)^2 dt. \tag{5}$$

Now we assume that the filtered function reaches its asymptote before the next step occurs, that is, the time constant τ is relatively small compared with the pulse duration α. For these *non-interfering* step responses, Eq. 5 reduces to:

$$\sigma_s^2 = \frac{W^2\alpha}{\beta} \qquad (\tau \ll \alpha). \tag{6}$$

Note that the asymptotic variance for non-interfering step functions is independent of the characteristic filter constant τ.

4.2.3 First Order Harmonic: at High Frequencies

When the temporal frequency $\omega = 1/2\beta$ is larger than the cut-off frequency ω_c, we may ignore the higher order harmonics of the modulation function $\delta(t)$ and restrict our analysis to the first order or fundamental spectral component. The amplitude of the fundamental component for a modulation amplitude W and pulse duration α is $4W\alpha\omega$ (for short pulse durations). This fundamental component is attenuated by the temporal filter $g(t)$ yielding a filtered amplitude $4W\alpha\omega\tilde{g}(\omega)$. The resulting variance $\sigma_f^2(\tau)$ of the filtered fundamental component is half its squared amplitude:

$$\sigma_f^2(\tau) = \frac{1}{2}\left(\frac{4W\alpha\omega}{1+(2\pi\tau\omega)^2}\right)^2. \tag{7}$$

The asymptotic behavior of $\sigma_f^2(\tau)$ (for a high frequency $\omega \rightarrow \infty$) is inversely proportional to τ^4:

$$\sigma_f^2(\tau) = \frac{1}{2}\left(\frac{W\alpha}{\pi^2\tau^2\omega}\right)^2. \tag{8}$$

4.3 Parameter Estimation Based on Pulse Duration Dependence

The dependence of modulation thresholds W as a function of the pulse duration α for a fixed pulse separation β is shown in Figure 3. This dependence allows to derive the two parameters that specify our model, that is, the time constant τ of the second order temporal filter $g(t)$, and the threshold variance σ_0^2 that determines the performance of the variance detection stage.

The time constant τ that characterizes $g(t)$ can be estimated using two data points of Fig. 3: the leftmost and the rightmost data point. We will use the data of subject HS to

illustrate the estimation procedure. The leftmost point shows the threshold $W = 1.2$ for pulse duration $\alpha = 14.3$ ms and pulse separation $\beta = 500$ ms. Such short (ideal) pulses filtering yields non-interfering impulse response functions. The threshold variance $\sigma_p^2(\tau)$ for non-interfering impulse responses is given by Eq. 4 (substitute $W = 1.2$, $\beta = 500$ ms, and $\alpha = 14.3$ ms). The variance expression $\sigma_p^2(\tau)$ depends on only one parameter: the time constant τ.

The rightmost point shows the threshold $W = 0.2$ for pulse duration $\alpha = 143$ ms and pulse separation $\beta = 500$ ms. Now, the pulse duration is 'finite' and the modulation function can be described by an incremental step followed by a decremental step. The variance σ_s^2 for these step responses is given by Eq. 5 (substitute $W = 0.2$, $\beta = 500$ ms, and $\alpha = 143$ ms). Because we had no a priori knowledge about the time constant τ, we have not used the asymptotic behavior (Eq. 6), but used the exact expression (Eq. 5).

The variances $\sigma_p^2(\tau)$ and σ_s^2 of the pulse and step modulation function are both equal to the threshold variance σ_0^2. Therefore, we may equate them: $\sigma_p^2(\tau) = \sigma_s^2$, and solve them numerically. Basically, this equation reflects the intersection of two asymptotes in the log-log presentation of W versus α. The first asymptote is the Bloch law for short pulse durations (with slope -1) (Gorea and Tyler, 1986). The second asymptote for large pulse durations is inherent to variance detection or probability summation, and has slope -1/2. Both asymptotes are shown in Fig. 3, based on the estimated threshold variance and time constant for each observer.

We find $\tau = 15 \pm 5$ ms (HS) and $\tau = 19 \pm 5$ ms (PW). Furthermore, Eq. 4 gives us the threshold variance σ_0^2 after substitution of the average time constant τ (and other parameters): $\sigma_0 = 9.9\%$ (HS) and 10.2% (PW).

4.3.1 Peak Detection?

One may wonder how critical the type of decision stage is for the estimation of the time constant τ of the second order low-pass filter. Although we found strong experimental support for variance detection (Werkhoven *et al.*, 1992), we will analyze another general class of decision mechanisms: peak detection. Contrary to variance detection (which is based on the square modulation amplitude), peak detection is based on the magnitude of the modulation amplitude. Peak detection monitors the maximum filter output for the modulation function. The modulation is detected whenever this maximum exceeds a threshold.

A peak detection model yields estimates for τ that are slightly higher ($\tau = 32$ ms for HS and 37 ms for PW) than estimated in a variance detection model. Although we observe a difference between the estimated time constants in a peak and a variance detection model, this is a minor discrepancy.

4.4 Parameter Estimation Based on Frequency Dependence

Above, we estimate the time constant τ based on the observed dependence of detection thresholds on pulse duration. In this section, we will test our variance detection model and estimate

τ based on the observed dependence of thresholds on the pulse separation β, or temporal frequency $\omega = 1/2\beta$. (see Fig. 4). We use the thresholds at two extrema: W at $\omega = 1$ Hz and W at $\omega = 12$ Hz. The reasoning (for this illustration for observer HS) is as follows.

At modulation frequency $\omega = 1$ Hz, the pulses (of duration $\alpha = 14.3$ ms) yield well-separated non-interfering impulse response functions after temporal filtering. The resulting variance $\sigma_p^2(\tau)$ for non-interfering impulse responses is given by Eq. 4 (substitute $W = 1.2$, $\alpha = 14.3$ ms, and $\beta = 500$ ms).

At frequency $\omega = 12$ Hz, we assume that the higher order spectral components of the modulation function are strongly reduced in amplitude by the temporal low-pass filter and may be ignored. Hence, we only consider the variance of the (attenuated) first order spectral (fundamental) component. The variance $\sigma_f^2(\tau)$ of the filtered fundamental component is given by Eq. 7 (with $W = 0.8$, $\alpha = 14.3$ ms and $\omega = 12$ Hz.

— Figure 6 about here —

Both variance expressions are dependent on only one parameter: time constant τ. The variance $\sigma_p^2(\tau)$ for ideal pulse functions and $\sigma_f^2(\tau)$ for the fundamental component are shown in Fig. 6 as a function of τ for HS (in fact we use the square root of variance, $\sigma_p(\tau)$ and $\sigma_f(\tau)$ are shown). Because the variances $\sigma_p^2(\tau)$ and $\sigma_f^2(\tau)$ of the pulse and fundamental modulation function are both equal to the threshold variance σ_0^2, we may equate them: $\sigma_p(\tau) = \sigma_f(\tau)$. From Fig. 6 we estimate the variances to be equal at time constant $\tau = 26 \pm 3$ ms. The threshold variance at this time constant is equal to the threshold standard deviation σ_0 and is 7.5%. Thus, the absolute detection modulation amplitudes are approximately 0.23 deg/s (HS) for the average speed $v_0 = 3$ deg/s used in this experiment.

For observer PW we estimated (using the same method): $\tau = 29 \pm 3$ ms and $\sigma_0 = 8.3\%$.

To substantiate our claim that higher order spectral components may be ignored for the computation of τ, we performed an analogous calculation using the 9 Hz threshold of Fig. 4 instead of the 12 Hz threshold. The computed τ for this case is $\tau = 24$ ms for observer HS, and is not significantly different (given a standard deviation of 3 ms). Therefore, 12 Hz may be considered far above the cut-off frequency.

The frequency dependence of threshold amplitude on frequency is characterized by two asymptotes. At low frequencies we have Bloch's law (see Eq. 4) and for high frequencies we consider only the fundamental component (see Eq. 8). Both asymptotes are shown in Fig. 4 based on the estimated threshold variance and time constant for each observer.

5 General Discussion

5.0.1 Evaluation

Werkhoven *et al.* (1992), proposed a two-stage detection model for velocity vector modulations (*i.e.*, modulations of both speed and direction): low-pass filtering followed by variance detection. This model has only two parameters: the characteristic time constant of the low-pass filter, and the threshold of the variance detection stage. Therefore, this model yields strong qualitative predictions when applied to the pulse detection experiments, that is, detection performance as a function of pulse duration and pulse separation. These predictions were tested.

First, a pulse duration α small compared to time constant τ yields impulse response functions $g(t)$ for each individual pulse. Moreover, when the pulse separation β is large compared to τ, these pulse responses do not interfere. Hence, at threshold, the amplitude of the response function is determined by the product of the pulse energy: the product of the pulse duration and the pulse amplitude: $W\alpha$. Pulses with equal energy result in detection. Therefore, we expect a hyperbolic relation for the dependence of threshold amplitude on pulse duration. This prediction (Bloch's law) is supported by the data (see Fig. 3. Second, at pulse durations α much larger than the time constant τ we expect that the variance of the filtered modulation function to be linear with pulse duration and quadratic with the threshold amplitude W. For this condition, thresholds are expected to deviate form Bloch's law and to be inversely proportional to the square root of pulse duration. This predicted deviation is observed in Fig. 3. Third, at high modulation frequencies, detection performance is expected to be dominated by the first order harmonic of the modulation function. For this condition, asymptotic detection thresholds increase linearly with modulation frequency (see Fig. 4).

The above observations strongly suggest that the detection of pulse-shaped modulation functions (studied here) and the detection of other functions described in Werkhoven *et al.* (1992) is carried out by closely similar, if not identical, mechanisms of the human visual system. One may wonder what constitutes this temporal filtering of the modulation function. Werkhoven *et al.* (1992) showed that the low-pass filter characteristics discussed above cannot be equivalent to an integration of the correlation output of standard motion detectors (Reichardt correlators). They showed that the temporal low-pass filter must be inherent to processing preceding the correlation stage. An analysis by Glünder (1990) strongly supports the view that a low-pass temporal filter stage corresponds with the inherently non-ideal band-pass *delay* filters in correlator detectors.

5.0.2 Characteristic Filter Time

The detection thresholds of pulse modulations of speed presented in this paper are modeled by two parameters: the characteristic filter time $\tau = 15 - 29$ ms and the variance detection threshold (square root variance $\sigma_0 = 8 - 10\%$).

The characteristic time constant of 15-29 ms corresponds to a temporal integration of

60-100 ms of the speed signal. As mentioned above, this integration time is determined by the temporal characteristics of the delay filter in Reichardt correlators (Glünder, 1990). An integration time of 60-100 ms is in excellent agreement with the upper temporal separation (inter stimulus interval) under which observers are still able to sense motion flow (Morgan and Ward, 1980).

The time constant ($\tau = 15 - 29$ ms), inferred from pulse modulation experiments presented in this paper, is lower than the estimate $\tau \approx$ 35-40 ms found for periodical triangular and sinusoidal modulation functions (Werkhoven *et al.*, 1992). This need not come as a surprise, since (as explained in the methods section) the present experiments are performed using dark targets against a photopic background, whereas the stimulus in Werkhoven *et al.* was a luminuous dot moving against a dark background. It is well known that the visual system becomes faster with increasing luminance (adaptation) levels (Kelly, 1971).

5.0.3 Variance Detection Threshold

Minimum speed modulation detection thresholds for the row of squares used in the present experiment (8% at 1 Hz) are markedly lower than the minimum thresholds (17% at 1 Hz) for the single blob target used in Werkhoven *et al.*. This improvement in detection performance is likely to result from the stimulus optimizations for this experiment versus Werkhoven *et al.*: (1) the number of targets simultaneously moving is four (versus one), (2) the observer was provided a uniform motion path as an explicit reference for detection (versus NO reference).

5.0.4 Speed Modulation Detection v. Speed Discrimination

In typical speed *discrimination* experiments observers view two uniform motion paths, separated in time or space, and indicate the path with the highest (or lowest) speed. Speed discrimination thresholds (the relative threshold difference between high and low speed) reported are as low as 6% (DeBruyn and Orban, 1988). These 6% *difference* thresholds suggest 'equivalent' *modulation* thresholds as low as 3% for spatiotemporal contiguous block-shaped motion paths. Perhaps surprisingly, the modulation detection thresholds (presented in this paper) for spatiotemporal block-shaped contiguous paths are 8%.

We suggest that this difference is caused by the *uncertainty* of the observer with respect to the phase of the modulation functions. This uncertainty may force the observer to use the autocorrelation (variance) of the velocity modulation signal, instead of the more efficient cross correlation of the modulation signal received with the signal expected (Green and Swets, 1966; Burgess and Ghandeharian, 1984).

5.0.5 Speed Modulations v. Direction Modulations

Bloch's law found for small pulse durations and large pulse separations revealed that, for these conditions, modulation detection is based on a spatial cue: the spatial excursion of the dot in the moving frame. This threshold spatial excursion is estimated approximately 3 arcmin for

observer HS and 3.5 arcmin for PW at the average speed of 3 deg/s and 1.6 deg eccentricity for this experiment.

Werkhoven *et al.* (1992) also estimated spatial excursion thresholds, but for modulations *orthogonal* to the average motion path. They found a smaller spatial excursion threshold of 2 arcmin at a lower average speed of 1.7 deg/s and larger eccentricity 3.4 deg. Note, that Werkhoven *et al.*'s stimulus was also less optimal because of the single target used and the absence of a reference motion path.

A comparison of the spatial excursion thresholds parallel and thresholds orthogonal to the motion path suggests that the visual motion system is more sensitive to modulations in motion direction than to modulations in motion speed.

5.0.6 The Invalidity of 'Window of Visibility' Theories

We have shown that speed modulation detection thresholds vary only slightly with the luminance contrast of the targets. This finding is at odds with 'window of visibility' theories (Watson *et al.*, 1986; Burr *et al.*, 1986) as we will argue below.

The 'Window of Visibility' theory was originally proposed to explain the discriminability of smooth and sampled (apparent) motion in terms of differences in 'visible' spectral components. Sampled motion has a Fourier transform that is the convolution of the spectrum for the unsampled (smooth) motion with a regular series of temporal pulses. The energy of certain components for the sampled motion (not present for smooth motion) within a 'visible' spectral window would perceptually distinguish sampled from smooth motion.

A similar theory for speed *modulations* would predict detection whenever the modulated motion path has detectable spectral components not present for uniform motion. The detectability of spectral components that result from speed modulation is determined by the modulation amplitude as well as by target luminance contrast. In fact, we expect a trade-off between threshold modulation amplitude and luminance contrast. This is not supported by the data that show similar thresholds for luminance contrasts of 25%, 50% and 100%.

To explain modulation detection, 'window of visibility' theories would have to be rephrased in terms of *discriminability* of spectral components instead of *detectability*.

5.0.7 Conclusions

Optic acceleration detectors are absent in the human visual system. The detection of variations in target speed is *not* based on the *temporal derivative* of the speed modulation signal. Instead, speed modulation detection is based on the *variance* of the (temporally blurred) modulation function.

Modulation detection performance is determined by two parameters: the characteristic time constant of the low-pass filter and the threshold of the variance detection stage. This simple model accurately describes detection performance of pulse-shaped speed modulation functions (the present paper), as well as the detection of periodic (triangular, sinusoidal and block-shaped) modulations of speed and direction (Werkhoven *et al.*, 1992).

6 Acknowledgement

The research of Peter Werkhoven was supported by the USAF Life Science Directorate, Visual Information Processing Grant 88-0140. Herman Snippe was supported by the InSight project of the ESPRIT Basic Research Actions of the European Community.

References

[1] Arnspang J. (1988) Optic Acceleration, *Proceedings ICCV*, 364–373, 1988.

[2] Bloch A.M. (1885) Expériences sur la vision, *C. R. Seances Soc. Biol. Paris, 37*,493-495.

[3] Burgess A and Ghandeharian H (1984) Visual signal detection. I. Ability to use phase information, *J. of the Optical Society America A* **1**, No. **8**, 900–905.

[4] Burr D.C., Ross J. and Morrone M.C. (1986) Smooth and sampled motion, *Vision Research* **26**, 643–652.

[5] DeBruyn B. and Orban G.A. (1988) Human velocity and direction discrimination measured with random dot patterns, *Vision Research* **28**, 1323–1335.

[6] Glünder H. (1990) Correlative velocity estimation: visual motion analysis, independent of object form, in arrays of velocity tuned bilocal detectors, *J. of the Optical Society America A* **7**, No. **2**, 255–263.

[7] Gorea A. and Tyler C.W. (1986) New look on Bloch's law for contrast, *J. of the Optical Society America A* **3**, No. **1**, 52–61.

[8] Green D.M. and Swets J.A. (1966) Signal detection theory and psychophysics, New York: Wiley.

[9] Kelly D.H. (1971) Theory of flicker and transient responses, I. Uniform fields, *J. of the Optical Society America* **61**, 537–546.

[10] Koenderink J.J. and Doorn A.J. van (1978) Detectability of power fluctuations of temporal visual noise, *Vision Research* **18**, 191–195.

[11] Koenderink J.J. (1986) Optic flow, *Vision Research* **26**, 161–180.

[12] Lange H. de (1958) Research into the dynamic nature of the human fovea-cortex system with intermittent and modulated light, *J. of the Optical Society America* **48**, 777-789.

[13] Morgan M.J. and Ward R. (1980) Conditions for motion flow in dynamic visual noise, *Vision Research* **20**, 431–435.

[14] Rashbass C. (1970) Unification of two contrasting models of the visual incr· nental threshold, *Vision Research* **16**, 1281–1283.

[15] Roufs J.A.J. (1972) Dynamic properties of vision-I. Experimental relationships between flicker and flash thresholds, *Vision Research* **12**, 261–278.

[16] Roufs J.A.J. (1973) Dynamic properties of vision-III. Twin flashes, single flashes and flicker fusion, *Vision Research, 13*, 309-323. *Vision Research* **13**, 309–323.

[17] Roufs J.A.J. (1974) Dynamic properties of vision-VI. Stochastic threshold fluctuations and their effect on flash-to-flicker sensitivity ratio, *Vision Research* **14**, 871–888.

[18] Watson A.B. (1979) Probability summation over time, *Vision Research* **19**, 515–522.

[19] Watson A.B., Ahumada A.J. and Farrell J.E. (1986) Window of visibility: a psychophysical theory of fidelity in time-sampled visual motion displays, *J. of the Optical Society America A* **3**, No. **3**, 300–307.

[20] Werkhoven P., Snippe H.P. and Toet A. (1992) Visual Processing of Optic Acceleration. submitted for publication

7 Figure Captions

Figure 1: A sketch of the spatial stimulus structure. The stimulus contains two rows of 4 square targets. The fixation mark is in the center of the configuration. The top and bottom row move (coherently) in opposite direction. One of them is the modulated motion path, the other the uniform path.

Figure 2: Speed modulation function $\delta(t)$. The function is characterized by alternating positive and negative pulses relative two the average speed v_0. The amplitude of the positive pulses is dv, that of the negative pulses $-dv$. The pulse duration is α. The pulse separation β is the time interval between the on-set of two consecutive pulses.

Figure 3: Speed modulation detection thresholds $W = dv/v_0$ as a function of pulse duration α. The pulse separation β was constant (500 ms). Note the rotated error bar of the rightmost data point, where pulse duration was varied at a fixed, 20%, modulation amplitude dv/v_0 to determine the threshold duration. The solid line (with slope -1) shows the Bloch law prediction ($W\alpha = constant$), that is, the asymptote for short pulse durations fitted through the leftmost data point. The dashed line is the asymptote for long pulse durations (see Model Section).

Figure 4: Speed modulation detection thresholds $W = dv/v_0$ as a function of the temporal modulation frequency $\omega = 1/(2\beta)$. Frequency ω is the frequency of the fundamental spectral component of function $\delta(t)$. Thresholds are given for a pulse duration $\alpha = 14.3$ ms and are estimated using threshold measured at $\alpha = 14.3, 28.6$ and 42.9 ms (see text for detailed explanation).
The graph is presented in log-log coordinates. The solid line with slope -1/2 shows asymptotic model predictions for low frequencies. The line with slope 1 shows asymptotic model predictions for high frequencies (see Model Section).

Figure 5: Speed modulation detection thresholds $W = dv/v_0$ as a function of target luminance contrast. Filled squares: Pulse duration was $\alpha = 28.6$ ms, pulse separation $\beta = 500$ ms ($\omega = 1$ Hz). Open squares: Pulse duration was $\alpha = 14.3$ ms, pulse separation $\beta = 56$ ms ($\omega = 9$ Hz).

Figure 6: Predicted variance as a function of characteristic time constant τ for two conditions for observer HS. Open squares: $\sigma_p^2(\tau)$ for ideal pulse functions (based on 1 Hz threshold modulation amplitude). Filled squares: $\sigma_f^2(\tau)$ for the fundamental component (based on 12 Hz threshold modulation amplitude).

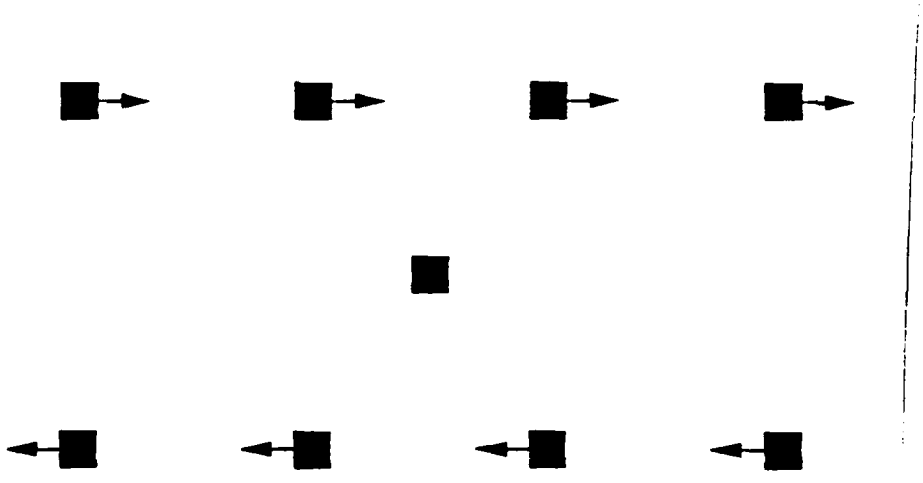


Figure 1: A sketch of the spatial stimulus structure. The stimulus contains two rows of 4 square targets. The fixation mark is in the center of the configuration. The top and bottom row move (coherently) in opposite direction. One of them is the modulated motion path, the other the uniform path.

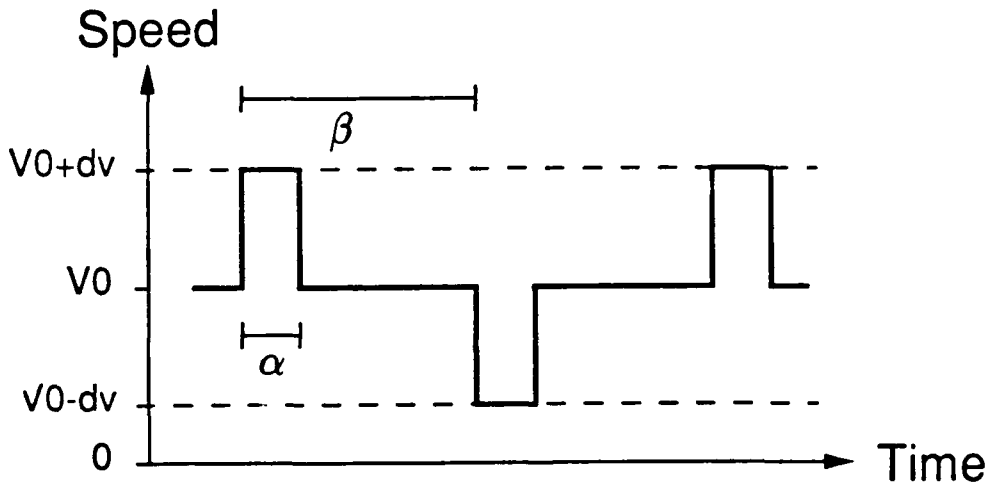


Figure 2: Speed modulation function $\delta(t)$. The function is characterized by alternating positive and negative pulses relative two the average speed v_0. The amplitude of the positive pulses is dv, that of the negative pulses $-dv$. The pulse duration is α. The pulse separation β is the time interval between the on-set of two consecutive pulses.

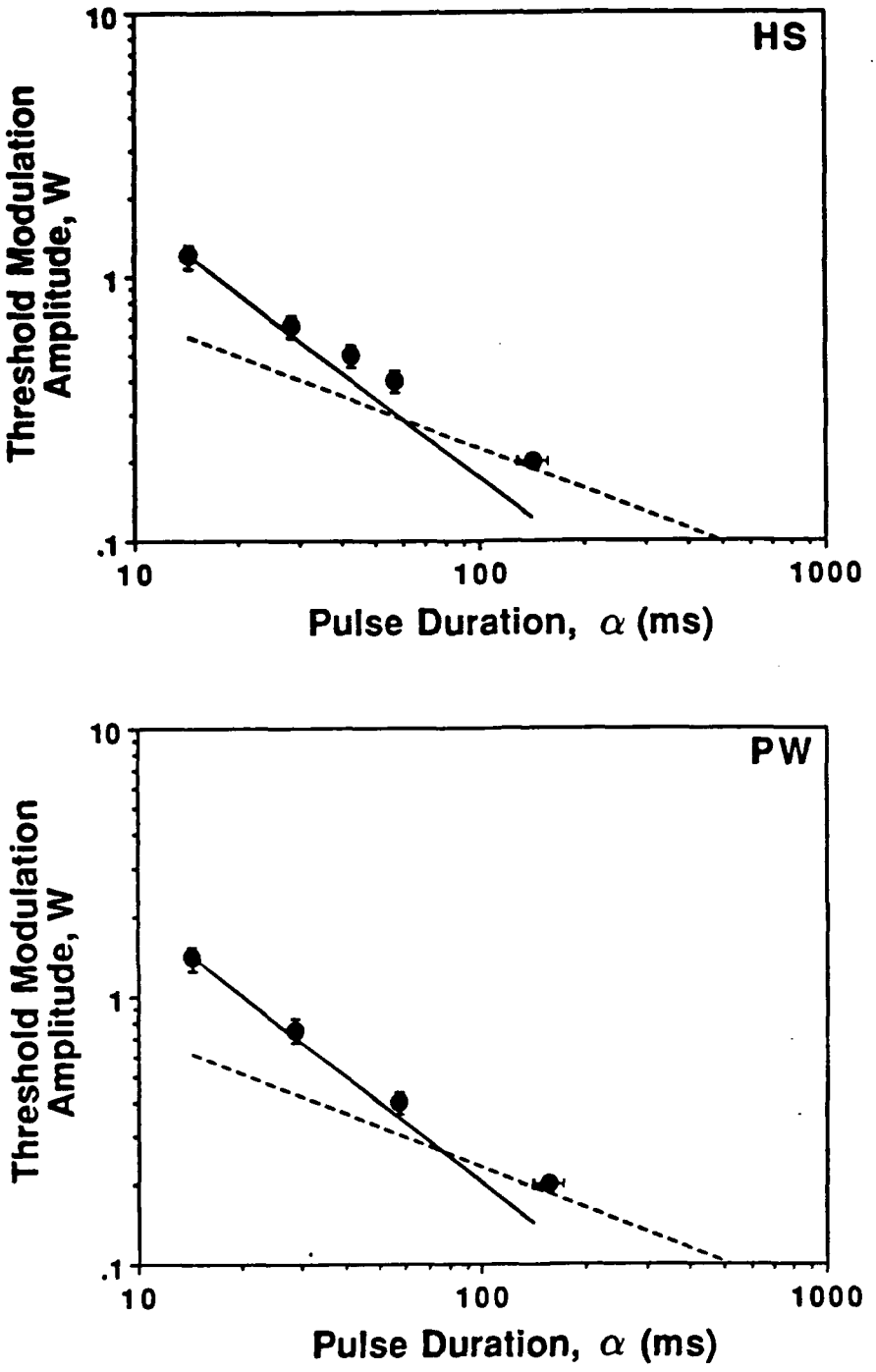
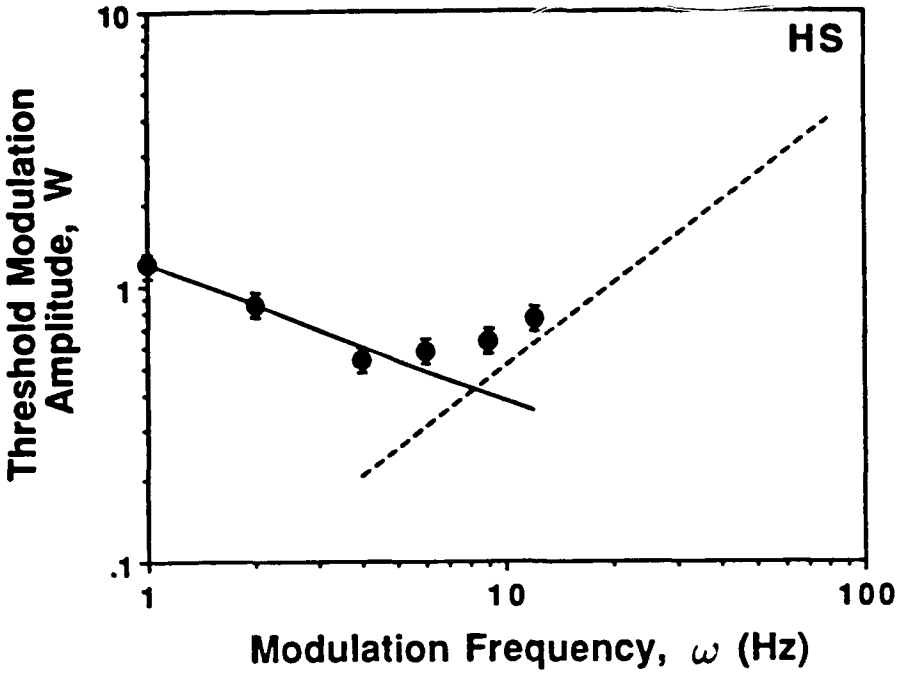


Figure 3: Speed modulation detection thresholds $W = dv/v_0$ as a function of pulse duration α. The pulse separation β was constant (500 ms). Note the rotated error bar of the rightmost data point, where pulse duration was varied at a fixed, 20%, modulation amplitude dv/v_0 to determine the threshold duration. The solid line (with slope -1) shows the Bloch law prediction ($W\alpha = constant$), that is, the asymptote for short pulse durations fitted through the leftmost data point. The dashed line is the asymptote for long pulse durations (see Model Section).



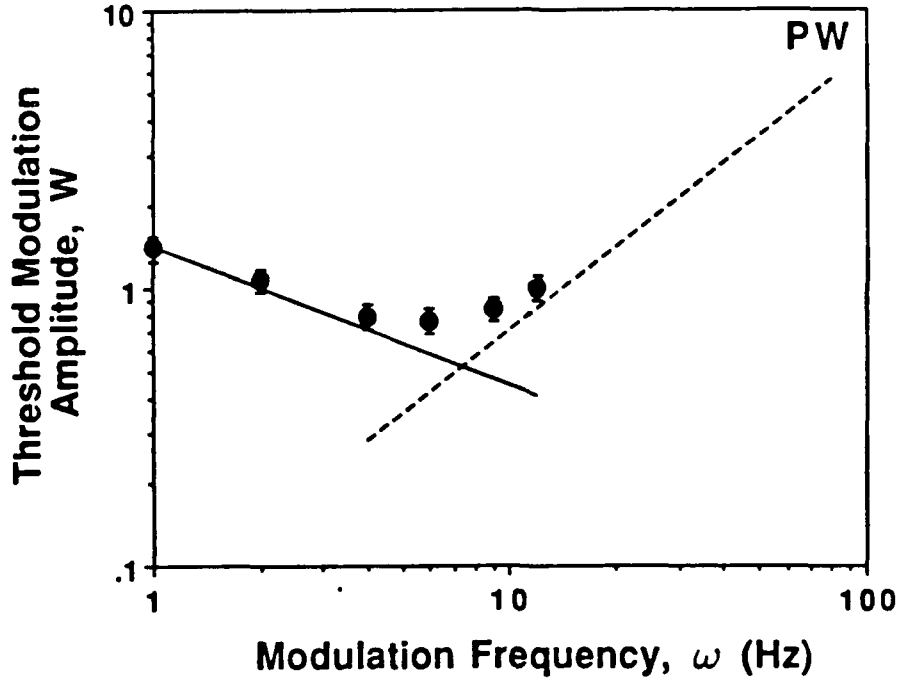


Figure 4: Speed modulation detection thresholds $W = dv/v_0$ as a function of the temporal modulation frequency $\omega = 1/(2\beta)$. Frequency ω is the frequency of the fundamental spectral component of function $\delta(t)$. Thresholds are given for a pulse duration $\alpha = 14.3$ ms and are estimated using threshold measured at $\alpha = 14.3, 28.6$ and 42.9 ms (see text for detailed explanation).

The graph is presented in log-log coordinates. The solid line with slope -1/2 shows asymptotic model predictions for low frequencies. The line with slope 1 shows asymptotic model predictions for high frequencies (see Model Section).

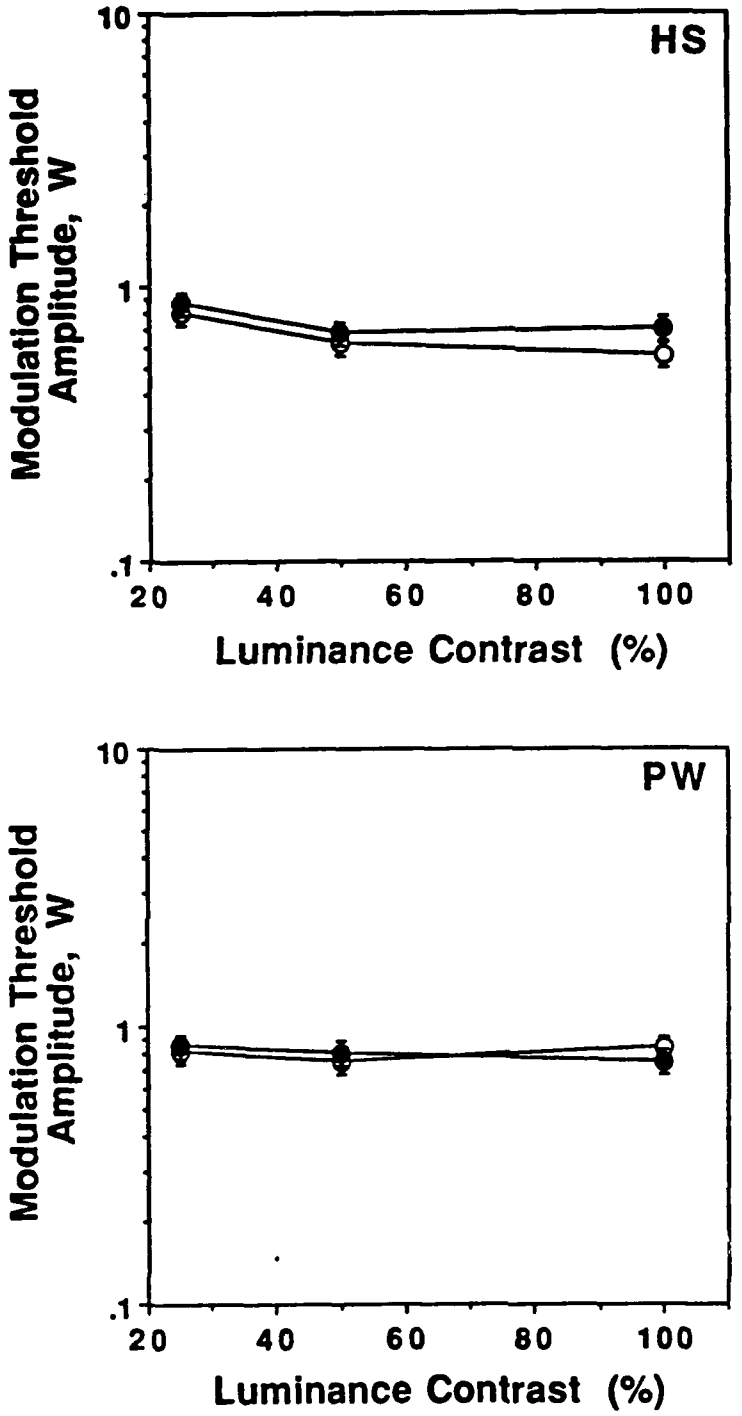


Figure 5: Speed modulation detection thresholds $W = dv/v_0$ as a function of target luminance contrast. Filled squares: Pulse duration was α = 28.6 ms, pulse separation β = 500 ms (ω = 1 Hz). Open squares: Pulse duration was α = 14.3 ms, pulse separation β = 56 ms (ω = 9 Hz).

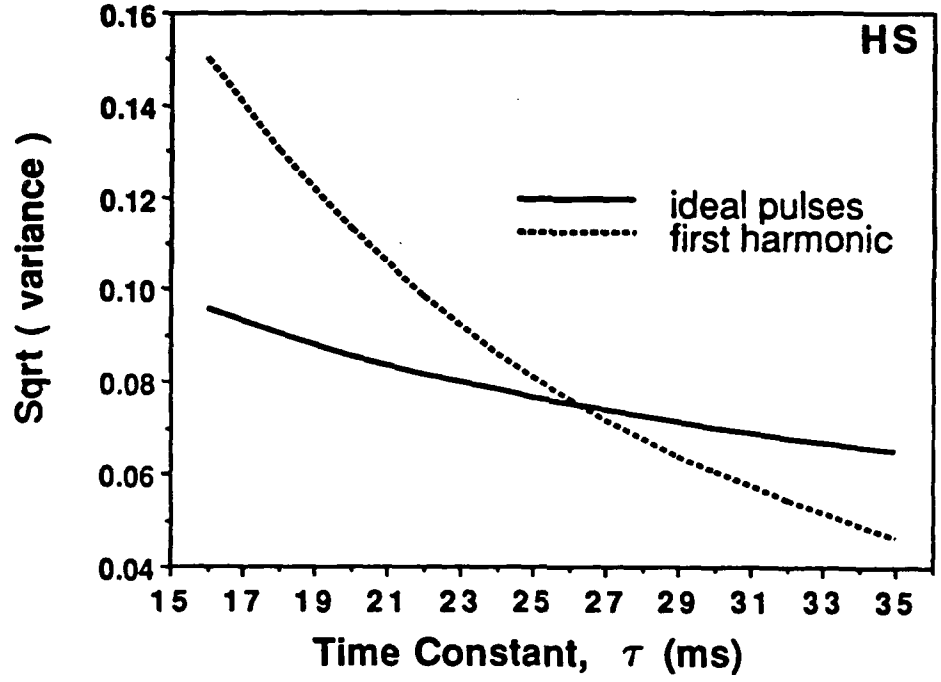


Figure 6: Predicted variance as a function of characteristic time constant τ for two conditions for observer HS. Open squares: $\sigma_p^2(\tau)$ for ideal pulse functions (based on 1 Hz threshold modulation amplitude). Filled squares: $\sigma_f^2(\tau)$ for the fundamental component (based on 12 Hz threshold modulation amplitude).